Niveau(x) d'enseignement: 3 Numéro EDUQ : 8170726
Centre d'intérêt: 72

Lamarche, Rolande, 1944-.

Le rôle des règles phono-graphiques dans l'acquisition de l'orthographe lexicale / Rolande Lamarche. -- Montréal : Université de Montréal, Programme de perfectionnement des maîtres de français au primaire, 1981.
x, 248 p. ; 28 cm. -- (Le Français à l'école primaire. Recherches et essais ; 4)

Bibliographie: p. 239-248. Glossaire: p. 237.

ISBN 2-920156-04-7

Cote Dewey: 372.6'32/L215r Cote LC: PC/2066

Descripteur(s): *Français (Langue) *Langue écrite *Orthographe *Test *Primaire (Niveau)

Identificateur(s) de sujet: *Rythme d'apprentissage

L'étude, qui porte sur l'acquisition de l'orthographe lexicale, vise deux objectifs: comprendre par quels mécanismes un individu acquiert la maîtrise de l'orthographe; découvrir s'il existe chez les enfants une intégration inconsciente du code phono-graphique du français.

Suite à une revue de la documentation, l'auteur présente une épreuve orthographique constituée de logatomes qu'il a soumise à un échantillon de 195 écoliers du primaire, dont 35 de deuxième année et 40 de chacun des niveaux de la troisième à la sixième année. L'échantillon provient de deux écoles dont les élèves sont de langue maternelle française et ne présentent aucun problème majeur d'apprentissage. Le niveau d'études et le rendement général en français ont été retenus comme variables indépendantes; la méthode d'apprentissage et le niveau socio-économique, comme constantes. L'expérience a eu lieu en classe, au cours du 3e mois de l'année scolaire, en l'absence du professeur.

La recherche a établi que l'acquisition des règles de formation des suites graphémiques correctes en français est un phénomène complexe. Tandis que certaines règles sont assez bien maîtrisées dès la 2e année, d'autres ne le sont encore que très mal à la fin du cours primaire. De plus, dans le cas de certaines règles, il semble exister un plafonnement que même les élèves forts ne peuvent dépasser, du moins au cours des six premières années de scolarité.

Type de document: r
Financement: Bénéficiaire - Université de Montréal
Organisme(s) source(s) - MEQ - DGEC
Programme(s) de subvention - PPMF
Auteur de l'analyse: gc

E 0 0 0 813028

Le rôle des règles phono-graphiques dans l'acquisition de l'orthographe lexicale

Rolande Lamarche

Recherches et essais
Numéro 4

Le français à l'école primaire

Université de Montréal
Faculté des sciences de l'éducation
P.P.M.F. primaire

E D U Q 8170726

LE FRANÇAIS À L'ÉCOLE PRIMAIRE
SÉRIE : RECHERCHES ET ESSAIS

NUMÉRO 4

LE RÔLE DES RÈGLES PHONO-GRAPHIQUES DANS L'ACQUISITION DE L'ORTHOGRAPHE LEXICALE

ROLANDE LAMARCHE

PROGRAMME DE PERFECTIONNEMENT DES MAÎTRES DE FRANÇAIS AU PRIMAIRE
UNIVERSITÉ DE MONTRÉAL

1981

Conception graphique de la page couverture: Daniel Bergeron

Cette recherche a été réalisée grâce à une subvention octroyée par le Ministère de l'éducation du Québec au Programme de perfectionnement des maîtres de français au primaire (P.P.M.F.) de l'Université de Montréal.

Dépôts légaux
Bibliothèque nationale du Québec
Bibliothèque nationale du Canada

ISBN 2-920156-04-7

© Centre de diffusion, Programme de perfectionnement des maîtres de français au primaire, Université de Montréal, 1981.

responsable du Centre de diffusion: Suzanne Fortin
Claude Germain
Elca Tarrab

AVANT-PROPOS

Ce texte est essentiellement celui de ma thèse de doctorat qui a été soutenue à l'Université de Montréal sous le titre: Intégration du code orthographique: aspects théoriques et vérification empirique. Le texte original a toutefois subi quelques modifications afin d'en rendre la lecture plus aisée.

Pour mener ce projet à terme, j'ai reçu des appuis financiers provenant d'abord du Conseil des Arts du Canada et du service de l'enseignement supérieur du Ministère de l'éducation du gouvernement du Québec et, plus tard, du Ministère de l'éducation du gouvernement du Québec via le programme de perfectionnement des maîtres de français au primaire de l'Université de Montréal.

Tout au long de l'élaboration de ce travail, j'ai bénéficié de l'expertise de nombreuses personnes dont celle de mon directeur de recherche, monsieur Gilles Gagné, ainsi que celle de madame Suzanne Winsberg et de messieurs Laurent Santerre et Richard Kittredge.

Enfin, pour la mise au point des versions préliminaire et finale de ce travail, j'ai pu compter sur l'aide de mesdames Carol Major, Céline Beauchemin, Monique Boissonneault et Michèle Coulombe.

TABLE DES MATIÈRES

LISTE DES TABLEAUX .....

INTRODUCTION 1

CHAPITRE I : L'APPRENTISSAGE DE L'ORTHOGRAPHE: PROBLÉMATIQUE 9

1.1 Introduction 9
1.2 L'orthographe française 12
1.3 Recherches pédagogiques sur l'orthographe 24
1.4 Rapport entre la lecture et l'écriture 30
1.5 L'orthographe anglaise 34
1.6 Recherches sur la lecture 37
1.7 Recherches sur l'apprentissage de la lecture . 48
1.8 L'apprentissage de l'orthographe lexicale: hypothèses 57

CHAPITRE II : MÉTHODOLOGIE 65

2.1 Introduction 65
2.2 Construction de l'épreuve orthographique 67
2.3 Caractéristiques de l'échantillon 89
2.4 Déroulement de l'expérience 99

CHAPITRE III : ANALYSE DES RÉSULTATS 103

3.1 Introduction 103

3.2 Vérification de l'hypothèse E1 105

3.2.1 L'indice de similitude au code français 106

3.2.2 Conclusion 120

3.3 Vérification de l'hypotèse E2 123

3.3.1 L'indice d'étendue du code 125

3.3.2 Les indices de popularité et d'utilisation des graphies 131

3.3.3 Conclusion 156

3.4 Vérification de l'hypothèse E3 163

3.4.1 Maîtrise de la règle R1 167

3.4.2 Maîtrise de la règle R2 174

3.4.3 Maîtrise de la règle R3 179

3.4.4 Maîtrise de la règle R4 183

3.4.5 Maîtrise de la règle R5 187

3.4.6 Maîtrise de la règle R6 191

3.4.7 Conclusion 195

3.5 Vérification de l'hypothèse E4 206

3.5.1 Le rapport fréquence empirique - fréquence théorique 208

3.5.2 Conclusion 217

CONCLUSION 221

GLOSSAIRE 237

BIBLIOGRAPHIE 239

LISTE DES TABLEAUX

2.01 Fréquences en français des expressions graphémiques du phonème /s/ en fonction de sa position dans un mot..... 78

2.02 Fréquences en français des expressions graphémiques du phonème /ɛ/ en fonction de sa position dans un mot..... 79

2.03 Fréquences en français des expressions graphémiques du phonème /o/ en fonction de sa position dans un mot..... 80

2.04 Nombre d'occurrences de chacun des environnements phonémiques du phonème /s/ conservés pour la construction de l'épreuve orthographique..... 85

2.05 Nombre d'occurrences de chacun des environnements phonémiques du phonème /ɛ/ conservés pour la construction de l'épreuve orthographique... 86

2.06 Nombre d'occurrences de chacun des environnements phonémiques du phonème /o/ conservés pour la construction de l'épreuve orthographique... 87

2.07 L'épreuve orthographique..... 88

2.08 Coefficients de corrélation entre le rendement général en français et les rendements dans les quatre savoirs et en mathématiques... 93

2.09 Coefficients de corrélation entre le rendement général en français et les résultats à l'examen ministériel de français..... 95

2.10 Coefficients de corrélation entre le rendement général en français et le quotient intellectuel..... 97

3.01 Indices de similitude au code français pour le phonème /s/ en fonction du niveau d'études et du rendement en français..... 109

3.02 Résultats de l'analyse de variance effectuée à partir de l'indice de similitude au code français pour le phonème /s/..... 111

3.03 Indices de similitude au code français pour le phonème /ɛ/ en fonction du niveau d'études et du rendement en français..... 113

3.04 Résultats de l'analyse de variance effectuée à partir de l'indice de similitude au code français pour le phonème /ɛ/..... 115

3.05 Indices de similitude au code français pour le phonème /o/ en fonction du niveau d'études et du rendement en français..... 117

3.06 Résultats de l'analyse de variance effectuée à partir de l'indice de similitude au code français pour le phonème /o/..... 119

3.07 Indice d'étendue du code présenté en fonction du niveau d'études et du rendement en français. 126

3.08 Pourcentage d'étudiants de chaque niveau d'études et de chaque catégorie de rendement en français qui ont utilisé les diverses expressions graphémiques du phonème /s/..... 132

3.09 Pourcentages d'utilisation des expressions graphémiques les plus fréquentes pour représenter le phonème /s/..... 137

3.10 Pourcentages d'étudiants de chaque niveau d'études et de chaque catégorie de rendement en français qui ont utilisé les diverses expressions graphémiques du phonème /ɛ/..... 142

3.11 Pourcentages d'utilisation des expressions graphémiques les plus fréquentes pour représenter le phonème /ɛ/..... 145

3.12 Pourcentages d'étudiants de chaque niveau et de chaque catégorie de rendement en français qui ont utilisé les diverses expressions graphémiques du phonème /o/..... 150

3.13 Pourcentages d'utilisation des expressions graphémiques les plus fréquentes pour représenter le phonème /o/..... 153

3.14 Taux moyens de réussite à la règle R1 en fonction du niveau d'études et du rendement en français..... 168

3.15 Résultats de l'analyse de variance effectuée à partir des résultats obtenus à la règle R1..... 170

3.16 Taux moyens de réussite à la règle R2 en fonction du niveau d'études et du rendement en français..... 175

3.17 Résultats de l'analyse de variance effectuée à partir des résultats obtenus à la règle R2..... 177

3.18 Taux moyens de réussite à la règle R3 en fonction du niveau d'études et du rendement en français..... 180

3.19 Résultats de l'analyse de variance effectuée à partir des résultats obtenus à la règle R3..... 181

3.20 Taux moyens de réussite à la règle R4 en fonction du niveau d'études et du rendement en français..... 184

3.21 Résultats de l'analyse de variance effectuée à partir des résultats obtenus à la règle R4..... 185

3.22 Taux moyens de réussite à la règle R5 en fonction du niveau d'études et du rendement en français..... 188

3.23 Résultats de l'analyse de variance effectuée à partir des résultats obtenus à la règle R5..... 190

3.24 Taux moyens de réussite de la règle R6 en fonction du niveau d'études et du rendement en français..... 192

3.25 Résultats de l'analyse de variance effectuée à partir des résultats obtenus à la règle R6..... 194

3.26 Comparaison des fréquences empiriques des graphies exigées pour la réussite des règles et des taux de réussite obtenus à ces règles..... 201

3.27 Rapports fréquence empirique - fréquence théorique pour le phonème /s/ en fonction du niveau d'études..... 211

3.28 Rapports fréquence empirique - fréquence théorique pour le phonème /ɛ/ en fonction du niveau d'études..... 215

INTRODUCTION

Toutes les grandes civilisations ont contraint les individus à intégrer certains codes nécessaires à leur insertion fonctionnelle dans la société. Ainsi en est-il du code orthographique. La conformité des écrits d'un individu aux normes établies en matière d'orthographe rend possible ou facilite la communication écrite en réduisant les ambiguïtés et en favorisant la vitesse de lecture. En effet, la conformité de l'orthographe d'un individu à la norme reconnue contribue à éliminer certaines ambiguïtés dans les messages à transmettre puisque l'orthographe permet, dans certains cas, de faire la distinction entre les mots homophones. Par ailleurs, le lecteur qui rencontre un mot mal orthographié doit interrompre momentanément sa lecture pour tenter de donner un sens au mot en question s'il croit avoir affaire à un mot qu'il ne connaît pas ou pour en rectifier

l'orthographe s'il devine quelle réalité il veut représenter.

L'importance de la maîtrise du code orthographique dans les sociétés occidentales est telle que lorsqu'elle ne sert pas tout simplement de critère à l'accession à un emploi, elle est utilisée comme indicateur du fonctionnement intellectuel de l'individu. Un employé qui ne peut se conformer à la norme orthographique peut en effet être soupçonné, par son patron, de ne pas posséder les habiletés intellectuelles qui justifient sa confiance et l'octroi de promotions. Plus généralement, une personne qui a une orthographe incorrecte peut voir son statut rabaissé par ceux qui lisent ses écrits. La maîtrise des règles orthographiques est ainsi un des critères importants utilisés socialement pour catégoriser les individus.

Cette exigence pour l'orthodoxie en matière d'écriture contraste avec le laxisme relatif dont jouit l'expression orale. Quelqu'un qui s'exprime oralement peut en effet se permettre des variantes phonétiques (prononciation, accent), syntaxiques ou même sémantiques sans que quiconque ne s'en scandalise. Que cette personne

commette quelques fautes d'orthographe dans un texte et elle trouvera beaucoup moins d'indulgence chez ses lecteurs.

Cette attitude a eu pour conséquence d'orienter des énergies importantes vers la correction orthographique. Ceux qui doivent communiquer par écrit consultent fréquemment dictionnaires et grammaires afin de s'assurer de la conformité de leur orthographe avec la norme reconnue. Cependant, c'est à l'école que la société a confié la responsabilité de rompre l'individu à la loi orthographique.

Ainsi, l'école primaire a longtemps été une institution centrée sur l'acquisition de la norme orthographique et de la grammaire de la langue écrite. Il semblait en effet impérieux que l'enfant sache, par exemple, quels mots ont un pluriel en *-oux*. C'est ainsi qu'on lui faisait mémoriser d'interminables listes de règles et d'exceptions à ces règles. Aujourd'hui, l'école tente d'adopter des méthodes d'apprentissage plus naturelles et plus vivantes. Le problème de l'apprentissage de l'orthographe demeure toutefois aigu puisque malgré le changement d'approche, l'orthographe française n'est toujours pas

mieux maîtrisée (Thimonnier, 1970, p.23; Beslais, 1965, p. 6). Une explication possible de cette situation résiderait dans la difficulté réelle d'intégration d'un code aussi complexe que le code français. Une autre explication possible découlerait du peu de connaissance que l'on possède sur les mécanismes d'intégration de l'orthographe.

Les travaux portant sur ce sujet sont en effet à peu près inexistants. Cette étude voudrait tenter d'apporter une contribution à ce champ de connaissances en s'intéressant particulièrement à l'acquisition de l'orthographe lexicale qui a pour objet la partie invariante des mots par opposition à l'orthographe grammaticale qui se rapporte aux règles relatives aux modifications grammaticales des mots.

Afin de comprendre par quels mécanismes un individu acquiert la maîtrise de l'orthographe lexicale, on peut chercher comment expliquer qu'un adulte francophone parvienne à écrire correctement des mots français qu'il ne connaît pas ou encore qu'un groupe de personnes puissent s'entendre sur les représentations graphémiques des mots

fictifs présentés comme des mots de la langue française (Lamarche, à paraître). Ces phénomènes sont en effet assez étonnants si l'on songe qu'en français, plusieurs phonèmes peuvent s'écrire de dizaines de façons. Le phonème /ɛ/ par exemple peut être représenté par plus de vingt expressions graphémiques (Burney, 1959). Aussi y a-t-il lieu de se demander si l'enfant qui fait l'apprentissage de l'orthographe lexicale n'apprendrait pas, malgré l'école, non pas l'orthographe de chacun des mots d'un lexique donné mais plutôt l'ensemble des règles de correspondance entre les suites phonologiques et les suites graphémiques qui constituent le code phono-graphique français. On comprendrait dans ce cas, qu'un individu qui maîtrise ces règles parvienne à écrire correctement des mots français qui lui sont totalement inconnus.

Essentiellement, la présente étude a pour objectif d'examiner s'il existe, chez les enfants qui apprennent l'orthographe lexicale française, une intégration inconsciente du code phono-graphique du français. Cette connaissance du code a été qualifiée d'inconsciente parce que plusieurs de ses règles ne sont pas enseignées et

ceux qui les maîtrisent ne sont souvent pas conscients de leur existence (Lamarche, à paraître).

Pour atteindre cet objectif, on s'arrête d'abord, dans un premier chapitre, sur quelques caractéristiques de l'orthographe française. On passe ensuite en revue quelques recherches qui ont porté sur l'apprentissage de l'orthographe. Comme on le fait remarquer, ces études ont peu contribué à augmenter les connaissances sur la façon dont se fait cet apprentissage. Les études qui ont porté sur l'apprentissage de la lecture ont par contre poussé beaucoup plus loin l'analyse des mécanismes d'apprentissage de cette habileté. Après avoir vu comment l'apprentissage de l'écriture s'apparente à celui de la lecture, on passe en revue les principales études qui ont été effectuées sur l'apprentissage de la lecture et on s'inspire de leurs résultats pour élaborer des hypothèses sur les mécanismes d'apprentissage de l'orthographe lexicale.

Afin de vérifier les hypothèses énoncées au premier chapitre, on construit une épreuve orthographique à laquelle est soumis un échantillon d'élèves de la deuxième

à la sixième année du cours primaire. Les étapes de la construction de cette épreuve orthographique ainsi que les caractéristiques retenues des sujets de l'échantillon sont présentées au deuxième chapitre.

Le troisième chapitre présente une analyse des résultats obtenus à l'épreuve orthographique. Pour discuter les hypothèses de travail, ces résultats sont mis en relation avec les caractéristiques retenues des sujets de l'échantillon.

Afin d'éviter toute confusion sur la définition de certains mots ou expressions auxquels on a recours dans ce texte, celui-ci est suivi d'un bref glossaire dans lequel apparaissent les définitions de ces mots et expressions tels qu'ils y sont utilisés.

CHAPITRE I

L'APPRENTISSAGE DE L'ORTHOGRAPHE: PROBLÉMATIQUE

1.1 Introduction

Saussure (1916) et plusieurs de ses contemporains (Sapir, 1921; Bloomfield, 1933) considéraient que la seule justification de l'écriture est de représenter la langue orale. Pour d'autres comme Uldall (1944) et Vachek (1959), l'écrit remplit des fonctions spécifiques qui sont différentes et complémentaires de celles de l'oral. Le débat sur la place de l'oral et de l'écrit dans la langue n'est pas terminé mais les deux ont pour rôle de représenter, de façon directe ou indirecte, la pensée. On peut ainsi dire qu'écrire, c'est représenter la pensée par des signes graphiques. Cette représentation peut se faire selon différents codes dont l'utilisation adéquate permet à ceux qui s'en servent de communiquer entre eux.

Les codes de représentation graphique de la pensée se rattachent, à divers degrés, à deux principaux types de systèmes, le système idéographique et le système graphémique ou alphabétique. Ce second système auquel appartiennent

la plupart des langues écrites du monde occidental représente la pensée par un système isomorphe aux langues orales. En effet, ces codes écrits ont été développés de façon à reproduire la double articulation de la langue orale. Aux monèmes oraux correspondent les monèmes écrits et aux phonèmes correspondent les graphèmes.

La relation de quasi biunivocité qui a existé à un moment donné entre les composantes de la langue orale et de la langue écrite correspondante est devenue beaucoup plus complexe dans certaines langues. Alors que des langues comme l'espagnol ont préservé la biunivocité entre les composantes de l'oral et de l'écrit en modifiant périodiquement l'orthographe pour lui faire suivre l'évolution de la langue orale, d'autres langues comme le français et l'anglais n'ont pas procédé à cet ajustement. Aussi la relation entre les phonèmes de leur langue parlée et les graphèmes de leur langue écrite est-elle complexe. On examinera, dans la section qui suit le niveau de complexité de l'orthographe française.

L'orthographe, ou façon d'écrire correctement selon un code, doit faire l'objet d'un apprentissage. Déjà, de nombreux chercheurs se sont penchés sur cette question. Une section ultérieure consiste en un exposé d'études qui ont déjà été faites sur l'apprentissage de l'orthographe lexicale française. Cet exposé est complété par un compte rendu

d'études effectuées sur les problèmes de l'apprentissage de la lecture puisqu'elles peuvent servir d'inspiration à l'étude de l'apprentissage de l'orthographe.

A la lumière des recherches effectuées sur l'orthographe lexicale française ainsi que sur les mécanismes d'apprentissage de la lecture, on formule, dans la dernière section du présent chapitre, des hypothèses de recherche sur l'apprentissage de l'orthographe qui, si elles étaient confirmées, pourraient conduire à l'élaboration de nouvelles stratégies d'enseignement de l'orthographe lexicale.

1.2 L'orthographe française

Si le français écrit représentait la pensée par un système isomorphe au français oral, il serait alors logique qu'à chaque constituant de base (phonème) de la langue parlée corresponde un constituant de base (graphème) de la langue écrite. Or, la situation est fort différente, et en 1960, Martinet déclarait:

> "En français, les différences entre la graphie et la phonie sont de nature telle que l'on peut dire, sans aucune exagération, que la structure de la langue écrite ne se confond pas avec celle de la langue parlée"
> (Martinet, 1960).

L'hiatus entre le français oral et le français écrit aurait été, selon Blanche-Benveniste et Chervel (1969), engendré par deux causes principales. D'abord, on a entrepris d'écrire le français avec les graphèmes du latin. Or leur nombre était inférieur au nombre de phonèmes du français. Deuxièmement, on a figé l'orthographe alors que le français parlé continuait à évoluer. En effet, c'est en 1835 que s'est terminé l'essentiel de l'histoire de l'orthographe française avec la sixième édition du Dictionnaire de l'Académie. Lors de la septième édition en 1879, on a bien adopté quelques modifications d'accents mais, essentiellement, l'orthographe française demeurait la même. La prononciation qu'elle traduit aujourd'hui serait toutefois, selon Burney (1959),

celle du XIIIe siècle. A cette époque, en effet, on écrivait, comme aujourd'hui, *rat, clerc, pas* mais on prononçait la consonne finale alors que maintenant cette consonne est muette. A ce sujet, Thimonnier écrit :

> "C'est en grande partie parce qu'elle correspond à une prononciation qui n'est pas la nôtre que notre orthographe nous paraît aujourd'hui tellement difficile"
> (Thimonnier, 1970, p. 17).

Ce caractère non phonétique de l'orthographe française existait cependant bien avant qu'elle ne soit fixée. Déjà en 1578, Rambaud affirmait :

> "Pour bien escrire faut auoir de lettres à suffisance : (...) la bouche pour se faire entendre a enuiron cinquante deux prononciations & la main n'a que dix huict signes"
> (Rambaud, 1578, p. 110-112).

En 1681, Veirasse d'Allais faisait une remarque semblable. Parlant de ceux qui les premiers ont écrit en français, il écrivait :

> "Ils firent servir une seule lettre pour représenter un, deux, trois & jusques à quatre sons différents ; & tout au contraire ils se servirent en certaines rencontres de deux, trois & quelquefois de quatre figures différentes pour exprimer un son très simple : & c'est là sans doute la principale source de désordre de notre Orthographe"
> (Veirasse d'Allais, 1681, cité dans Blanche-B. et Chervel, 1969, p. 54).

Outre la non correspondance en français entre les nombres de graphèmes et de phonèmes, plusieurs autres caractéristiques de notre orthographe ont été dénoncées comme étant au moins partiellement responsables des difficultés d'apprentissage qu'elle présente.

Comme le rapporte Burney (1959), avant que l'on ait recours aux accents, dans l'orthographe française, l'ouverture du *e* était généralement marquée par le redoublement de la consonne qui suivait le phonème /ɛ/: (ex : *appelle*). En 1740, l'Académie introduit l'accent grave mais celui-ci ne marquera l'ouverture du *e* que dans les mots où le phonème /ɛ/ n'est pas suivi par une consonne double. Ainsi, on écrit *jette* mais *achète*, *nette* mais *complète*.

De même, l'accent circonflexe avait originellement pour fonction de marquer l'allongement de la voyelle qu'il surmontait. Cet allongement était indiqué, avant 1740, par des lettres muettes comme dans *roole* pour *rôle*, *maistre* pour *maître*, *anme* pour *âne*, *voulte* pour *voûte*. Quand on introduisit l'accent circonflexe, il semble qu'on n'ait pas toujours tenu compte du fait que les lettres muettes pouvaient avoir un autre rôle que celui d'allonger les voyelles qu'elles précédaient. C'est ainsi que certaines voyelles brèves et ouvertes portent un accent circonflexe comme dans *hôpital*. Par ailleurs, certaines voyelles longues et fermées comme dans *axiome* et *zone* n'ont pas reçu l'accent

circonflexe.

Quelques auteurs dont Blanche-Benveniste et Chervel (1969) et Thimonnier (1970) ont cependant réussi à démontrer que le désordre orthographique n'est en partie qu'apparent et qu'une étude approfondie du phénomène permettrait de dégager des structures rationnelles dans l'orthographe française. Alors que Blanche-Benveniste et Chervel exploitent surtout la relation phono-graphique, Thimmonier rassemble les mots en séries analogiques dont il établit la cohérence orthographique.

Blanche-Benveniste et Chervel définissent le code phono-graphique comme "l'ensemble des règles de correspondance entre graphèmes et phonèmes pratiquées par un individu, un milieu ou une époque" (Blanche-Benveniste et Chervel, 1969, p. 55). Ils y distinguent trois étages. Le premier est composé des graphèmes simples auxquels on associe des phonèmes (ex.: *b* = /b/). Puisque notre langue compte beaucoup plus de phonèmes que de graphèmes, seule une partie des phonèmes est représentée par les graphèmes simples. Pour compléter le système graphique français, il existe un second étage constitué de couples de graphèmes qu'on appelle "digrammes" et qui sont affectés, chacun, à un seul phonème (ex.: *an* = /ã/). Une partie des phonèmes pour lesquels il n'existe pas de graphèmes simples est ainsi représentée par

les digrammes. Enfin, un troisième étage est constitué par les trigrammes qui sont des unités formées de trois graphèmes représentant un phonème (ex : *eau* = /o/).

Blanche-Benveniste et Chervel notent que, selon sa position dans le mot, un graphème peut représenter divers phonèmes. Ce n'est toutefois pas le cas de tous les graphèmes dont certains n'ont qu'une valeur de base. Celle-ci est définie comme la valeur liée au minimum de contraintes. La valeur de base de c par exemple est /k/ et non /s/ puisque c ne vaut /s/ que lorsqu'il est suivi de *e*, *i* et *y* alors que sa valeur /k/ est possible dans toutes les autres distributions.

En plus de leur valeur de base, certains graphèmes peuvent aussi avoir une valeur dite de position. Ils peuvent avoir cette valeur à cause de la place (initiale, médiane, finale) qu'ils occupent dans un mot. Par exemple, le x de *six* vaut /s/ parce qu'il est placé en position finale. Cette valeur de position est aussi fonction du graphème voisin. c dans *cire* vaut /s/ parce qu'il est suivi de *i*.

Quand un graphème n'a pas de valeur phonétique propre mais que sa présence modifie la valeur phonétique d'un graphème voisin, on dit qu'il a une valeur auxiliaire.

C'est le cas du *a* de *gain* qui deviendrait un mot différent, en l'occurrence *gin*, si le *a* n'y était pas. Un graphème peut encore se fondre à un digramme quand, en se combinant à un autre, il perd comme lui toute valeur propre, l'ensemble fonctionnant comme un graphème unique dont la valeur phonétique est indépendante de celle de ses composants (ex : *au* = /o/).

Enfin un graphème peut prendre la valeur zéro quand il est muet et qu'il n'affecte pas la valeur des graphèmes qui l'entourent (ex : le *p* de *coup*) (Blanche-Benveniste et Chervel, 1969 ; Piacère, 1974).

Ainsi, la langue française, si elle contient un plus grand nombre de phonèmes que de graphèmes simples a, par contre, développé tout un ensemble de règles plus ou moins complexes qui permettent de représenter chacun des phonèmes.

Le travail de Thimonnier a, lui aussi, consisté à rechercher les règles de l'orthographe française mais avec un point de départ différent. Selon cet auteur :

> "Il faut d'abord se référer à une analyse objective du système. Partant de là, on peut aisément distinguer les formes normales des anormales, dénombrer avec rigueur les anomalies réelles de l'écriture..."
> (Thimonnier, 1970, p. 8).

L'"analyse objective du système" dont il parle l'amène d'abord à découvrir l'existence de constantes phonético-graphiques qui consistent en un certain nombre de règles qui déterminent les rapports généraux entre les sons et les signes. Ces règles s'appuient sur une nouvelle théorie de la syllabation graphique qui peut être résumée par les deux principes suivants :

> P1: "Le mot écrit contient autant de syllabes que de sons voyelles (*ou-bli*, *ha-ri-cot*) ou de diphtongues (*i-noul*, *bien-tôt*)
>
> P2: "Les consonnes intérieures sont réparties de telle sorte que chaque diphtongue et chaque son voyelle ne puissent être précédés que d'une seule consonne ou de son équivalent, le groupe combiné. (*per-te*. *pat-tre*, *poul-pe*, *sou-ple*, *af-fli-gé*, *obs-ta-cle*, *obs-truc-tion*)"
> (Thimonnier, 1970, p. 23).

Cette analyse du système orthographique français lui permet ensuite d'observer que, sauf pour 5% des mots du dictionnaire, tous les autres peuvent être groupés en séries analogiques, c'est-à-dire en groupes de mots "caractérisés par un élément morphologique (radical, préfixe ou suffixe) possédant à la fois le même sens, la même valeur phonique et la même forme graphique" (Thimonnier, 1970, p.24). Ainsi Thimonnier observe que la consonne qui suit l' *a* privatif reste toujours simple d'où la série préfixale

constituée de mots tels *amoral*, *anormal*, *asymétrique* etc. De la même façon, il fera remarquer que les quarante mots de la famille de *terre* ont un *r* doublé d'où la série : *terrain*, *enterrer*, *territoire*, etc.

La recherche de Thimonnier l'amènera enfin à constater que pour un nombre restreint de mots (5%), des séries analogiques tendent à se confondre avec d'autres. Il appellera ces séries, les séries homonymes. Dans la plupart des cas, ces séries homonymes se décomposent en séries analogiques et peuvent alors être l'objet de règles comme les autres séries analogiques dont nous parlions plus haut. Ainsi, pour la famille de *honneur*, "seuls s'écrivent par un 'n' simple les mots formés sur le radical '*honor*'. D'où *honnête* mais *honorable*, *déshonneur* mais *déshonorer*, etc." (Thimonnier, 1970, p. 25). Cette analyse de l'orthographe française conduit finalement Thimonnier à conclure que seulement 228 mots dont 69 d'usage courant font montre d'une orthographe qui ne répond pas aux critères qu'il a établis. Il ressort donc, du travail de Thimonnier, que malgré sa grande complexité, l'orthographe française serait tout de même assez bien organisée.

1.2.1 Les mouvements de réforme de l'orthographe française

L'orthographe française a toutefois été considérée pendant longtemps et, malgré le travail de Thimonnier, encore aujourd'hui, comme un code plutôt désordonné. De plus, sa complexité et le manque d'économie ont amené certains esprits à vouloir la réformer comme l'ont fait d'ailleurs plusieurs autres communautés linguistiques. En effet, pendant que l'orthographe française demeure inchangée, plusieurs ont été réformées au cours du XXe siècle. La Chine, La Turquie, l'Espagne, le Portugal, la Hollande, la Russie, l'Allemagne, le Danemark, la Norvège, pour ne nommer que quelques pays, ont tous apporté des modifications souvent très substantielles à leur orthographe.

Malgré la dissemblance des langues et la diversité des problèmes qui se posaient, il est intéressant de remarquer que, abstraction faite des solutions apportées aux problèmes propres à chaque langue, les réformes réalisées dans ces différents pays portent généralement sur les mêmes points, soit : le dédoublement de lettres géminées quand elles se prononcent simples ; la suppression de lettres muettes étymologiques ; une tendance générale à l'élimination des digrammes grecs (*ph*, *th*, *ch*), une volonté d'adapter les mots savants et les autres mots d'emprunt à la phonétique, à l'accentuation et à la graphie nationales. (Burney, 1959, p.115)

L'orthographe française est demeurée inchangée mais ce n'est certainement pas dû à une absence de propositions de réforme. Celles-ci ont en effet été nombreuses et diversifiées. On peut, pour les circonscrire, les classer en quatre groupes (Burney, 1959).

Le premier comprend les mouvements de réforme pour une orthographe libre. Ceux-ci sont apparus au XVIe siècle alors qu'on commençait à vouloir imposer une orthographe française unique. Les tenants de ces mouvements s'y opposaient en soutenant que l'utilisation de l'écriture en général et des lettres muettes en particulier doit refléter les états d'âme de chacun et ne peut donc pas être encadrée par des règles fixes.

Les mouvements de réforme pour une orthographe phonétique constituent le second groupe. Ces réformateurs voulaient que l'écriture soit une transcription fidèle de l'expression orale de chacun. Chaque dialecte ou même chaque variante aurait donc son orthographe. Ainsi à chaque réalisation d'un phonème correspondrait un graphème.

Au XVIIe siècle, les Précieuses réagissent devant la complexité de la langue écrite et militent en faveur de l'abolition "des lettres superflues". Au moment de la Révolution, même Napoléon qui, dit-on, n'aurait pas connu l'orthographe, se serait opposé au "mensonge abécédaire". En 1901, la Commission des Tolérances propose des

"tolérances" qui ne sont acceptées par l'Académie que dans les cas où elles concernent des règles spécifiques et sans grande conséquence. En 1904, une Commission de Simplification est nommée. Ses conclusions, soumises à l'Académie, sont presque toutes refusées. En 1939, Dauzat soumet de nouvelles propositions mais essuie lui aussi un échec. Il sera suivi, en 1944, par la Commission Langevin et en 1960, par Beslais. A ce troisième groupe de réformateurs prudents de l'orthographe, on peut ajouter Thimonnier pour qui, à part quelques exceptions, l'orthographe française formerait un système complexe mais cohérent, "régi par quelques principes fort simples, pratiquement applicables à tous les cas, et notamment à ceux qui, de prime abord, semblent décourager toute tentative de réglementation : familles de mots irréguliers, accentuation, consonnes géminées, lettres polyvalentes etc." (Thimonnier, 1970, p. 267).

Enfin, un quatrième groupe réunit les mouvements pour une orthographe phonologique. Les adeptes de ces mouvements voudraient qu'à chacun des phonèmes d'une langue qui aura dû être standardisée, ne corresponde qu'un seul graphème. Déjà au XVIe siècle, Meigret et Ronsard préconisaient une telle orientation pour l'orthographe française. C'est dans cette lignée que s'inscrivent les propositions de Blanche-Benveniste et Chervel.

Quelle que soit la direction qu'ont adoptée les différents mouvements de réforme de l'orthographe française, il se sont tous heurtés à une ferme opposition. Il ne reste donc plus au pédagogue que de prendre son parti de la situation et de développer des méthodes d'apprentissage susceptibles de faciliter l'assimilation d'un code aussi complexe.

1.3 Recherches pédagogiques sur l'orthographe

Il ressort, des travaux de Thimonnier, Blanche-Benveniste et Chervel que l'orthographe française est fort complexe. Pourtant, tout francophone doit en faire l'apprentissage. Il n'est donc pas étonnant que de nombreux pédagogues se soient penchés sur le problème.

C'est vers le milieu du XIXe siècle, a-t-on mentionné précédemment, que l'orthographe française a été fixée. On a d'abord perçu le problème de son apprentissage uniquement en terme de délimitation de ce qu'un étudiant doit apprendre. Un demi-siècle plus tard, on se rendait toutefois compte, grâce à Binet (1904) qu'il fallait aussi prendre en considération ce que l'écolier est en mesure d'apprendre. Mais, en parlant de cette époque, Guion, note que:

> "l'importance d'une génétique du vocabulaire et de l'orthographe n'est pas encore soupçonnée, pas plus qu'on ne met en doute l'idée plus ou moins inconsciente suivant laquelle apprendre l'orthographe, c'est apprendre tous les mots de la langue française c'est-à-dire tous ceux qui sont dans les dictionnaires" (Guion, 1974, p. 114).

Les choses évoluent quand même relativement rapidement puisque déjà en 1926, une étude de la Société Binet se propose de déterminer à quel âge les différentes règles de grammaire sont acquises (Pellet, 1926, p. 52).

En 1935, Buyse précise la problématique de l'orthographe en percevant la nécessité de délimiter la matière à enseigner, le meilleur moment pour le faire et la méthode à utiliser pour l'enseigner. Il voit aussi l'importance de développer des instruments qui puissent mesurer les résultats obtenus.

Quatre recherches seront élaborées à partir de cette problématique. La première est celle d'Aristizabal (1938) qui tentera d'établir une liste des mots qu'un enfant doit savoir écrire au terme de son cours primaire. Le corpus qu'il établira contiendra surtout des mots propres au vocabulaire de l'enfant sans toutefois exclure celui de l'adulte. Il comprendra 4 329 mots qui représentent 95% des 12 038 mots relevés dans les textes dépouillés.

La seconde étude est celle de Dubois (1939). Se basant sur le travail d'Aristizabal, Dubois se fixe comme objectif de construire un barème de difficulté orthographique et de déterminer les bases d'un programme de cours pour l'étude systématique de l'orthographe usuelle.

> "Le service que nous voulons rendre à l'école primaire par notre recherche, c'est de fournir aux maîtres des normes des difficultés telles qu'elles apparaissent dans les écrits des enfants aux âges scolaires successifs" (Dubois, 1939, p. 10).

Dubois construira donc, à partir du vocabulaire fondamental établi par Aristizabal et qu'il aura dicté à un grand nombre d'enfants, une échelle "dans laquelle les mots de difficulté identique ou voisine formeront des ensembles qui seront eux-mêmes distants l'un de l'autre d'un degré de difficulté mathématiquement équivalent" (Dubois, 1939, p. 84). Cette échelle sera construite à partir du pourcentage moyen de réussite de chaque mot par l'ensemble des enfants qui constituaient son échantillon.

Au terme de son travail, Dubois note qu'il serait fort utile de connaître quelle est la difficulté spécifique de chaque mot ou, en d'autres termes, "quelle est la faute la plus souvent commise quand on écrit ce mot. ... Il est en effet souhaitable que les maîtres soient avertis de la difficulté qui, pour tel mot, tient habituellement les élèves en échec" (Dubois, 1939, p. 100). Ce travail d'analyse des fautes sera effectué par Lambert (1947) qui tentera d'établir, pour chaque mot, un indice de difficulté. Pour établir cet indice, il fera l'inventaire des cacographies pour chaque mot du vocabulaire de base d'Aristizabal et calculera aussi, pour chacun d'eux, le nombre d'erreurs commises par les écoliers sur chaque partie du mot. Il observera, entre autres choses, comme le rapporte Guion, que le nombre de fautes de même que celui des cacographies diminuent au cours de la scolarité.

> "Dans les premières années, le nombre de cacographies trouvées passe par son maximum : on peut parler alors d'une dispersion maximale. Puis, au fur et à mesure que s'affermit la maîtrise graphique du mot, la dispersion diminue et l'on ne rencontre plus que quelques cacographies typiques"
> (Guion, 1973, p. 132).

Pirenne (1949) fera la synthèse des travaux effectués par Aristizabal, Dubois et Lambert et tentera de répondre à l'un des problèmes de l'apprentissage de l'orthographe qu'avait énoncé Buyse concernant la question de la répartition du programme au cours des années de scolarité. Ceci constituera la quatrième étude inspirée par la problématique formulée par Buyse. Après avoir étudié plusieurs critères de classement des mots constituant le vocabulaire de base, il en retient finalement deux, soit la fréquence d'emploi par les enfants ou l'intérêt et la difficulté orthographique. Au sujet de ces critères, il écrit :

> "Il semble qu'il faille accorder au facteur *intérêt* une importance plus grande qu'à la *difficulté*. Celle-ci ne devrait intervenir que lorsque son degré est tel qu'il réduit considérablement les chances de succès de l'apprentissage d'un mot au niveau scolaire auquel l'intérêt demande qu'on l'attribue"
> (Pirenne, 1949, p. 42).

Le vocabulaire de base sur lequel ont travaillé Aristizabal (1938), Dubois (1939), Lambert (1947), et Pirenne (1948) sera retouché, corrigé et enrichi par Ters, Mayer et Reichenbach et sera connu sous le nom de: Vocabulaire Orthographique de base

qui comprend 7961 mots (Ters et al., 1977). Ces mots devraient recouvrir, de l'avis de Guion, environ 97% des occurrences de n'importe quelle page de texte français non exagérément littéraire ou technique (Guion, 1974, p. 142).

Deux études plus récentes méritent d'être signalées. Il s'agit des études d'El Wakil (1969) et de Massarenti (1970) (in Guion 1973). Celle d'El Wakil, en plus de mesurer l'acquis orthographique des élèves, calcule des indices qui permettent de distinguer les mots faciles des plus difficiles et propose des exercices portant sur des mots difficiles. Quant à l'étude de Massarenti, elle se classe, pour ce qui concerne l'étude du problème de l'acquisition de l'orthographe lexicale, dans la même catégorie que les études belges conduites par Buyse puisqu'elle établit, elle aussi, un vocabulaire de base et une échelle de difficulté orthographique.

Les études qui viennent d'être mentionnées sur l'apprentissage de l'orthographe semblent toutes postuler que la maîtrise d'un certain vocabulaire de base doit résulter de l'acquisition systématique de l'orthographe de chacun des mots de ce vocabulaire. On constate que le relevé des difficultés spécifiques de chaque mot, qui représente un travail monstre, n'a conduit qu'au calcul des indices de difficulté des mots et qu'à leur classification en termes de mots plus ou moins faciles ou difficiles. Jamais, semble-t-il, on a regroupé les erreurs pour voir s'il ne

pourrait pas, par exemple, en ressortir des informations sur la façon dont se ferait l'apprentissage de l'orthographe lexicale.

Ces études qui fournissent des vocabulaires de base et indiquent les endroits, à l'intérieur des mots, où les erreurs sont le plus souvent commises ne renseignent donc pas sur la façon dont se fait l'apprentissage de l'orthographe lexicale. Elles ne font que supposer que celle-ci est acquise par la mémorisation de chacun des mots de la langue et cherchent des moyens pour favoriser une mémorisation efficace en tenant compte de variables telles l'intérêt ou la difficulté de chacun des mots.

Les études qui ont porté sur l'apprentissage de la lecture ont par contre poussé beaucoup plus loin l'analyse des mécanismes d'apprentissage de cette habileté. Avant de passer en revue les principales d'entre elles, on verra d'abord comment l'apprentissage de l'orthographe s'apparente à celui de la lecture.

1.4 Rapports entre la lecture et l'écriture

La lecture et l'écriture sont deux opérations qui associent des suites graphémiques et leurs représentations mentales. Ce trait commun n'implique pas cependant que leurs caractéristiques principales soient identiques. Il y a en effet entre ces deux opérations, des différences appréciables. L'activité de lecture en est une de décodage ou de reconnaissance alors que l'activité d'écriture consiste à encoder ou à générer un message. Or, la reconnaissance d'un mot ou d'un objet, fera remarquer Smith (1971) nécessite beaucoup moins d'informations que sa reproduction. A la suite d'une étude de Shannon (1951), on sait en effet que même quand, dans un texte, une lettre sur deux a été omise, le lecteur parvient malgré tout à comprendre et sans grande difficulté. Par contre celui qui écrit doit reproduire fidèlement chacun des éléments des mots qu'il écrit. Ceci correspond d'ailleurs à la remarque que Goodman fait au sujet de la lecture quand il écrit :

> "La lecture efficace ne relève pas de la perception précise et de l'identification de tous les éléments, mais de l'habileté à choisir les repères qui se révèlent exacts à la première tentative"
> (Goodman, 1967, p.2).

De plus les symboles écrits, comme le faisait remarquer Cronnell (1971), sont concrets et peuvent être décrits en tant qu'unités physiques alors que les stimuli de

l'écriture sont beaucoup plus imprécis. C'est peut-être ce qui peut expliquer que les enfants réussissent beaucoup plus tôt à différencier la plupart des lettres qu'à décomposer les mots en leurs différents sons (Liberman, Shankweiler, Carter et Fisher, 1972). Morf faisait une remarque semblable lorsqu'il écrivait :

> "Alors qu'à 6 ans, l'enfant aurait atteint sur le plan visuel une capacité d'analyse, il n'en est pas de même sur le plan auditif. Ses perceptions auditives sont encore au stade du syncrétisme ; il est incapable de situer ses perceptions auditives les unes par rapport aux autres dans une succession temporelle. Il n'est pas non plus capable d'une activité d'analyse d'un ensemble de sons"
> (Morf, 1966, p. 31).

Selon Smith (1973), le lecteur est avantagé par rapport à celui qui écrit du fait que son information va de la structure de surface d'un symbole écrit vers la structure profonde du sens (Chomsky et Halle, 1968). Pour celui qui écrit, les mêmes avantages n'existent pas puisque pour lui, la direction de l'information est inversée. C'est ainsi que le lecteur peut tirer profit des nombreuses redondances orthographiques, syntaxiques et sémantiques et n'est pas obligé d'identifier chacun des mots ou même de connaître leur signification. Pour celui qui écrit, par contre, non seulement ne peut-il pas profiter de ces redondances mais de plus, les contraintes qui lui sont imposées sont beaucoup plus rigides puisque l'écriture ne souffre généralement

aucune variation.

Comme fera remarquer Venezky:

> "The few sound-to-spelling rules that a reader might use are generally not bidirectional - the diagraph *ph* is usually pronounced 'f' but it would be risky to conclude that 'f' is usually written *ph*. Instead an independent set of sound-to-spelling rules would seem to be required by the writer" (Venezky, 1967).

On notera également que la lecture est facilitée par la médiation sémantique. En effet, selon Goodman (1971) et Smith (1971), le contexte dans lequel se trouve le lecteur ainsi que les mots qui entourent l'élément lexical à décoder fournissent beaucoup d'indices sur cet élément. On peut généralement comprendre une phrase dans laquelle un mot est absent. Les contextes linguistiques et extra-linguistiques donnent en effet suffisamment d'indices pour que le lecteur puisse deviner le mot manquant.

Enfin, à cause du nombre souvent plus restreint de phonèmes que d'expressions graphémiques dans une langue, l'opération de décodage qu'est la lecture présente un moins grand éventail de réalisations phonologiques que l'opération d'encodage qu'est l'écriture ne présente de réalisations

graphiques. Par exemple, la lecture du graphème é n'offre qu'une possibilité de réalisation phonologique alors que la transcription graphique du phonème /ɛ/ offre plus de vingt possibilités de réalisations graphiques. Il arrive par contre que la lecture soit plus complexe que l'écriture. En effet, alors que le phonème /ø/ ne peut s'écrire que d'une façon, il peut être lu comme un /ə/ ou un /ɛ/.

A cause des nombreuses différences entre la lecture et l'écriture, les conclusions des études portant sur l'apprentissage de la lecture ne peuvent conséquemment pas être transposées sans réserve à celui de l'écriture. Vu toutefois la rareté des études portant sur la façon dont se fait l'apprentissage de l'écriture, les conclusions des études portant sur la lecture pourraient peut-être contribuer à orienter le choix des hypothèses de travail dans le domaine de l'apprentissage de l'orthographe et aider à identifier, s'il en existe, d'autres facteurs d'apprentissage que la mémorisation. Ces études sont cependant, pour la plupart, des études américaines. On est toutefois peut-être justifié de s'y référer puisque, comme on le verra dans la section qui suit, les orthographes française et anglaise possèdent de nombreux points de ressemblance.

1.5. L'orthographe anglaise

Les anglophones, qui se sont souvent fait reprocher, comme les francophones, le niveau de désordre de leur orthographe, ont entrepris des études qui leur ont fait découvrir, comme l'a fait Thimonnier pour la langue française, que leur langue possède un niveau de cohérence relativement satisfaisant.

Chomsky soutient en effet que l'orthographe anglaise est "... a near optimal system for the lexical representation of English words". (Chomsky, N. et al. 1968). Quant à Carol Chomsky, elle a tenté de démontrer que contrairement à ce qu'on pourrait croire, plusieurs des éléments de l'orthographe anglaise ne sont pas arbitraires mais correspondent à un niveau de représentation, à l'intérieur du système phonologique de la langue, qui est plus profond que le niveau phonétique (Chomsky, C. 1970). Venezky et Weir (1966), pour leur part, se sont appliqués à démontrer que si l'on cherche au-delà d'une relation biunivoque entre graphèmes et phonèmes, on trouve une grande correspondance entre la langue anglaise écrite et celle qui est parlée.

Non seulement l'orthographe anglaise, comme l'orthographe française, n'est pas strictement phonémique mais elle possède de plus, comme l'orthographe française, un code

phono-graphique "à étages". Gillooly écrit en effet, au sujet de la langue anglaise:

> "... In a system such as English, the graphemic units are more complex, involving not only single letters but also compounds (letter sequences) formed of more than one letter. Examples are the digraphs ‹th›, ‹ch›, ‹ck›, etc. that function in a way different from the simple, combined effects of their constituent letters. (...) As a consequence of these graphemic patterns, English orthography employs far more than 26 functional units". (Gillooly, 1973, p. 171).

Alors qu'un groupe de recherche a identifié soixante-neuf ensembles graphiques qui jouent chacun le rôle d'un graphème simple dans le système graphique anglais (Berdiansky, Cronnel et Koehler, 1969), l'analyse de Venezky (1967) en a extrait soixante-cinq. Il ressort également de cette dernière étude, l'existence, dans le système graphique anglais, de "marker units" que Gillooly définit comme suit:

> "These are letters or letter sequences which, although sometimes being unsounded themselves, affect the sound of other letters in their environment (as the ‹e› in mate affects the vowel sound of ‹a›)". (Gillooly, 1973, p. 171).

Cette citation n'est certes pas sans rappeler les travaux de Blanche-Benveniste et Chervel sur les valeurs possibles des graphèmes dans le code phono-graphique français. On peut donc conclure que les orthographes anglaise et fran-

çaise ont plusieurs traits communs. Elles sont en effet toutes deux non phonétiques. Elles ont de plus chacune un éventail de digrammes qui jouent le rôle de graphèmes simples pour représenter des phonèmes. Elles ont aussi un ensemble de règles qui régissent la valeur des graphèmes, valeur qui peut varier en fonction de l'environnement.

A cause de ces nombreux points de ressemblance, ces deux orthographes présentent possiblement des problèmes d'apprentissage analogues que les personnes confrontées à ces difficultés solutionneraient de façon relativement similaire. Aussi, on passera en revue les principales études américaines qui ont été effectuées sur l'apprentissage de la lecture. Il est en effet peut-être possible d'utiliser les résultats de ces études pour comprendre comment les francophones franchissent les diverses étapes de cet apprentissage. Il pourrait en découler des hypothèses concernant l'apprentissage de l'orthographe et, plus précisément, de l'orthographe lexicale française.

1.6 Recherches sur la lecture

La lecture est l'opération mentale qui consiste à associer, à des signes graphiques, certaines représentations mentales. D'après Smith (1971), la lecture permet d'associer à des signes graphiques trois éléments différents qui sont la lettre, le mot et l'idée. Ces trois types d'associations correspondent à trois tâches distinctes qui peuvent être exécutées indépendamment l'une de l'autre à partir du même matériel.

Afin de parvenir à élaborer des méthodes d'apprentissage plus efficaces, de nombreux chercheurs ont tenté de découvrir les stratégies dont se servent les lecteurs pour en arriver à extraire le sens d'un texte écrit. En particulier, on a cherché à établir l'importance des tâches qui consistent à associer une suite graphémique à la représentation mentale des graphèmes qui la composent ainsi qu'à la représentation mentale des mots qui la contiennent. Dans les études effectuées sur les langues occidentales, on a aussi tenté d'établir le rôle de la médiation phonétique dans la lecture puisque les systèmes d'écriture occidentaux sont basés sur une représentation des idées isomorphe à l'expression orale de ces idées. Les critères utilisés pour apprécier les performances en lecture sont l'exactitude des associations entre les suites graphémiques et leur représentation mentale ainsi que la rapidité avec

laquelle ces associations sont établies. Puisque la finalité de la lecture est la communication des idées, les chercheurs évaluent les aptitudes et les stratégies en fonction de ce critère ultime.

Il a semblé naturel, puisque le mot écrit est composé d'une suite de graphèmes, de se demander si le mot n'est pas lu par la perception individuelle des graphèmes qui le constituent. La lettre serait alors l'unité de base de la lecture.

Selon Smith (1971), l'identification d'une lettre est réussie après que le lecteur a pu identifier ses traits caractéristiques. Cependant, il croit.comme Neisser (1967), que l'identification d'un mot n'est pas nécessairement le résultat de l'identification de chacune des lettres qui le constituent. Il existerait plutôt des traits distinctifs propres aux mots et que le lecteur utiliserait pour les identifier. Dans ce modèle, la reconnaissance d'une lettre à l'intérieur d'un mot exigerait l'identification d'un moins grand nombre de traits caractéristiques que sa reconnaissance lorsqu'elle est isolée.

Selon Gibson et al.(1962), l'unité de perception en lecture ne peut pas être la lettre puisque celle-ci n'a pas de correspondant acoustique unique en anglais. L'unité de perception ne peut pas non plus être le mot puisque,

toujours selon ces auteurs, comme unité de perception, le mot n'offrirait aucune base à partir de laquelle de nouvelles combinaisons graphiques pourraient être décodées et constituerait donc une unité qui ne coïnciderait pas avec un processus économique. Par ailleurs, Cattell (1885) a démontré que lorsqu'un mot n'est ni trop long ni trop rare, il peut être perçu, par un lecteur expérimenté, aussi rapidement qu'une simple lettre. L'unité de perception serait donc plus étendue que le graphème.

Par ailleurs, l'aptitude à associer des phonèmes aux graphèmes ne serait pas une condition suffisante à la lecture. L'orthographe finlandaise est probablement la plus régulière en usage aujourd'hui. Or Venezky (1973) a conduit, auprès d'enfants finlandais des trois premières années du cours primaire, une étude qui consistait à demander aux sujets de lire une liste de mots synthétiques. Les résultats indiquent que pour les étudiants de deuxième et de troisième années, la corrélation entre l'habileté à associer les phonèmes aux graphèmes et l'habileté en lecture est plutôt faible. Ces résultats ont donc amené Venezky à conclure que " ... the ability to generalize letter-sound correspondences did not guarantee high ability in reading" (Venezky, 1973, p. 288).

Pour établir le rôle de la médiation phonétique dans la lecture, Gibson et son équipe (1962, 1963) ont effectué

une série d'expériences dans lesquelles on compare l'aptitude à identifier des suites de graphèmes qui constituent des mots de l'anglais, des mots fictifs prononçables en anglais et enfin, des mots difficilement prononçables dans cette langue. Les mots de chacune de ces catégories étaient composés des mêmes lettres; par exemple, *pig*, *gip*, *gpi*.

Les résultats indiquent que le lecteur expérimenté lit tout aussi facilement les mots fictifs prononçables que les mots de l'anglais. Par contre, les mots difficilement prononçables sont moins bien lus que les mots prononçables, surtout s'ils sont longs. Pour expliquer ce phénomène, les auteurs de l'étude ont émis une hypothèse qui peut être traduite de la façon suivante:

HYPOTHÈSE L1 : L'unité de perception dans la lecture, est constituée d'une suite de graphèmes qui entretient une relation invariante avec une suite donnée de phonèmes dans une langue déterminée.

Certains mots qui ont été utilisés par Gibson et al. dans leurs expériences et qui sont difficilement prononçables par les Anglais peuvent en fait être plus facilement prononçables par des individus d'autres langues. Si les Anglais ont de la difficulté à les prononcer, c'est qu'ils ne sont

pas habitués à associer ces suites de graphèmes à des suites précises de phonèmes. Il semble donc que la fréquence d'une suite de graphèmes dans les mots anglais devrait faciliter la lecture ou la perception de ces suites graphémiques.

Plusieurs études ont démontré qu'un mot est reconnu d'autant plus rapidement qu'il est fréquent. Des travaux de Howes et Solomon (1951), Solomon et Postman (1952), McGennies, Comer et Lacey (1952), Postman et Conger (1954), King-Ellison et Jenkins (1954), Baker et Feldman (1956), Moscovici et Humbert (1960) ont en effet tous confirmé ceci.

Postman et Rosenzweig (1956) ont, quant à eux, conduit une étude dans laquelle ils ont voulu comparer le seuil de reconnaissance de trigrammes fréquents parmi lesquels se trouvaient des mots anglais de trois lettres et celui de trigrammes rares. Ils constatent que les mots anglais de trois lettres ne sont pas reconnus plus rapidement que les syllabes familières constituées également de trois lettres. Par contre, les trigrammes rares exigent un temps de reconnaissance plus long.

On déduit de cette expérience que la rapidité de perception d'un mot n'est pas influencée par sa fréquence mais par celle d'une autre unité qui peut coïncider avec un mot

lorsque celui-ci est court.

L'hypothèse de Gibson et al. sur le rôle des relations invariantes entre suites de graphèmes et suites de phonèmes a été remise en question quand Gibson, Shurcliff et Yonas (1970) ont repris les expériences de Gibson et al. mais cette fois auprès d'enfants sourds. Les résultats obtenus n'étant pas significativement différents de ceux obtenus par les sujets entendants, Gibson et al. en ont conclu que l'unité de perception, en lecture, serait indépendante de son corrélat phonétique. Les auteurs de cette étude sont ainsi amenés à penser qu'il doit exister "a kind of grammar for letter sequences that generates permissible combinations without regard to sound" (Gibson et al. 1970). Cette hypothèse de Gibson et al. sur la lecture peut être traduite de la façon suivante :

HYPOTHÈSE L2 : Il existe dans la langue écrite des règles qui déterminent l'acceptabilité de suites graphémiques et les lecteurs expérimentés dans cette langue connaissent ces règles au moins implicitement.

Il n'est pas sûr cependant que l'on puisse conclure, avec Gibson et al., que la véracité de cette seconde hypothèse infirme celle de la première. Le fait que les lecteurs connaissent les règles de l'orthographe lexicale ne

s'oppose en effet pas à ce qu'ils lisent en ayant également recours à la médiation phonétique.

Les règles orthographiques impliquent qu'il y a une certaine redondance dans les suites graphémiques. En effet, s'il existe des règles, à une lettre donnée ne peuvent être attachées que certaines lettres de sorte qu'en lisant la première lettre, on a déjà des indications sur celles qui vont suivre. Par exemple, en français, en voyant *q*, on s'attend à voir à sa suite *u* ou un blanc. L'influence joue aussi dans l'autre direction. En identifiant un *h* qui n'est pas en début de mot, on déduit qu'il y a de fortes chances que la lettre précédente soit un *c*.

Plusieurs études ont confirmé les conclusions de Gibson et al. quant à la relation entre l'habileté à utiliser les redondances orthographiques et la compétence en lecture. A la suite d'une étude expérimentale dans laquelle on mesurait le temps de reconnaissance de mots fictifs dont le degré de ressemblance avec de véritables mots anglais variait, Wallach écrivait: "Good spellers were found to recognize nonsense words which resemble English much more readily than poor spellers". (Wallach, 1963, p. 61). Cette conclusion vient renchérir la proposition selon laquelle les étudiants qui réussissent bien en lecture, contrairement

à ceux qui réussissent moins bien, n'apprendraient pas seulement l'orthographe de mots pris séparément mais intégreraient également la structure de probabilités qui régit l'arrangement des lettres à l'intérieur de ces mots.

Costermans, à la suite d'une étude impliquant une analyse factorielle de la perception des mots écrit:

> "... The principal source of variance in word perception was S's ability to use his previous knowledge of the stimuli in reconstructing fragments as an effect of redundancy" (Costermans, 1967, p. 407).

En 1968, Wallace entreprend une étude semblable à celle de Wallach (1963). Son expérimentation consiste à demander aux sujets de choisir, parmi des couples de mots fictifs, celui qui ressemble le plus à un mot anglais. Le nombre de réponses correctes s'est avéré significativement plus élevé chez les bons lecteurs que chez les autres. Cette différence entre les bons et les mauvais étudiants en lecture ressortait avec plus d'évidence dans les cas des choix modérément difficiles que dans ceux qui étaient situés aux extrêmes.

Mason (1975) a, quant à elle, vérifié que la redondance spatiale, c'est-à-dire la fréquence de position de certaines lettres ou séquences de lettres, constitue un élément qui contribue à faciliter l'identification des lettres et donc

la lecture. Ses résultats indiquent de plus que les bons lecteurs utilisent mieux que les autres les indices fournis par la redondance spatiale pour identifier les lettres.

Les études conduites par Smith l'amènent également à souligner l'importance des redondances orthographiques dans la lecture. Parlant du lecteur, il écrit:

> "... if he acquires the knowledge that certain patterns of features just do not occur together in English words - then he can construct criterial subsets within the feature lists specifying word categories, and identify words on much less visual information."
> (Smith, 1971, p. 220).

Selon Goodman (1971), quand un lecteur est placé devant un mot qui lui est inconnu, il a d'abord recours à une stratégie basée sur la recherche de la correspondance graphème-phonème pour décoder le mot en question. Ses recherches l'ont toutefois amené à conclure que quand un individu lit un texte composé de mots connus, il utilise un nombre minimal d'indices visuels qu'il sélectionne en fonction de ses attentes. A mesure que ces indices sont traités mentalement, le lecteur prend des décisions que des indices sémantiques et syntaxico-sémantiques contenus dans la suite du texte lui feront conserver, rejeter ou raffiner. Il écrit:

> "Readers utilize ... three kinds of information

> simultaneously. Certainly, without graphic input, there would be no reading. But, the reader uses syntactic and semantic information as well. He predicts and anticipates on the basis of this information, sampling from the print just enough to confirm his guess of what's coming, to cue more semantic and syntactic information."
> (Goodman, 1971, p. 266).

Tout en ne s'opposant pas aux théories élaborées par les études citées précédemment qui s'intéressent surtout à l'identification de l'unité de perception en lecture, Goodman atténue l'importance des facteurs visuels dans la lecture au profit de la médiation sémantique et syntaxico-sémantique. Sur ce point, il rejoint Smith qui reconnaît, lui aussi, à ce niveau d'indice, un rôle important.

Puisque la finalité de la lecture et de l'écriture est la communication d'un contenu psychique, les travaux de Goodman et Smith qui portent davantage sur le décodage du sens d'un texte sont intéressants. Toutefois, dans le contexte de la présente étude, on se réfère aux recherches effectuées sur la lecture parce que leurs conclusions peuvent constituer un biais valable par lequel il serait possible d'élaborer des hypothèses sur les mécanismes d'apprentissage de l'orthographe lexicale, les études sur ce sujet étant à peu près inexistantes. Or, les règles de l'orthographe lexicale ne sont pas fondées sur le sens des mots. Il ne semble donc pas approprié de tenir compte, pour les fins de cette étude, des considérations portant sur le décodage du sens

d'un texte. On retiendra cependant que les conclusions de Smith et de Goodman n'excluent pas le rôle des indices visuels dans la lecture.

On peut donc conclure à partir de l'ensemble des études effectuées sur la lecture, que les deux hypothèses énoncées plus haut sont plausibles. En effet l'unité de perception, en lecture, pourrait être constituée, chez les entendants, d'une suite de graphèmes qui entretient une relation invariante avec une suite donnée de phonèmes dans une langue déterminée. Par ailleurs, les lecteurs expérimentés dans une langue connaîtraient, au moins implicitement, les règles qui déterminent l'acceptabilité de suites graphémiques.

Ces deux hypothèses étant vérifiées chez les lecteurs expérimentés, il reste à voir comment les apprentis-lecteurs intègrent ces habiletés. La section qui suit présente des études qui portent sur leur acquisition.

1.7 Recherches sur l'apprentissage de la lecture

Les bons lecteurs auraient développé deux habiletés fondamentales qui leur permettent, d'une part, d'associer à des suites graphémiques les suites phonologiques correspondantes et, d'autre part, d'identifier les suites graphémiques correctes dans une langue. Certaines études donnent des indications sur l'acquisition de ces habiletés par les apprentis-lecteurs.

1.7.1 Vérification de l'hypothèse L1 chez les apprentis-lecteurs

Pour expliquer comment les règles de correspondance de suites graphémiques à des suites phonologiques s'acquièrent lors de l'apprentissage de la lecture (Hypothèse L1), Gibson, Osser et Pick (1963) ont émis deux hypothèses. La première suppose qu'il y aurait assimilation des règles grâce à l'intervention du maître qui les ferait ressortir des mots que l'enfant aurait appris à lire de façon globale. Dans ce cas, en début d'apprentissage, les règles seraient encore inconnues et l'enfant ne pourrait pas décoder des mots fictifs. La seconde hypothèse suppose que l'apprenti-lecteur assimilerait ces règles de la même façon qu'il assimile la grammaire de sa langue parlée. Aussitôt qu'il comprendrait que l'écriture représente des sons, il commencerait à rechercher les constantes du langage écrit.

Pour vérifier laquelle de ces hypothèses correspond à la réalité, Gibson, Osser et Pick ont mis au point une expérience qui les a amenés à constater que dans les premiers stades de l'apprentissage de la lecture (première année), l'enfant lit par petites unités mais aurait déjà commencé à généraliser certaines règles de correspondance de suites graphémiques à des suites phonologiques puisqu'il peut lire correctement certains mots fictifs. Au cours des stades plus avancés (troisième année), l'enfant lit par unité plus longue et se réfère clairement à des règles de correspondance de suites graphémiques à des suites phonologiques. Des règles de correspondance seraient donc perçues dès les premiers stades de l'apprentissage de la lecture.

Si ce modèle d'apprentissage correspond à la réalité, l'enfant devrait apprendre à lire plus rapidement dans une langue dans laquelle les règles de correspondance graphème-phonème sont simples et les unités entre lesquelles s'établissent des relations invariantes, petites, c'est-à-dire qu'à un seul graphème correspond toujours un même phonème. Or, plusieurs études susceptibles de trancher la question ont déjà été effectuées.

Ruddel a comparé le rendement en lecture de jeunes Américains de première année dont certains avaient fait leur apprentissage de la lecture par un alphabet phonétique

(Initial Teaching Alphabet) et les autres par l'alphabet anglais traditionnel. Les résultats l'amènent à la conclusion suivante :

> "The first-grade reading programs possessing a high degree of consistency in grapheme-phoneme correspondences produced significantly higher a) word reading, b) word study skills and c) regular word identification achievement than the reading programs making little provision for consistent correspondences with the exception of one contrast on the word study skills variable" (Ruddel, 1968, p. 267).

L'allemand qui avait été classé par Bloomfield (1942) comme étant une langue dans laquelle la correspondance entre graphèmes et phonèmes est relativement régulière a souvent fait l'objet de comparaison avec l'anglais. Plusieurs conclusions ressortent de ces études. Parmi celles-ci, mentionnons celle de Preston (1953) et celle de Samuels (1969) qui démontrent que les enfants allemands de première et de deuxième années sont nettement supérieurs aux enfants américains des mêmes niveaux dans des tâches de reconnaissance de mots.

En 1963, Preston entreprend une nouvelle étude qui démontre cette fois que les enfants américains de quatrième et de sixième années lisent beaucoup plus vite que les enfants allemands d'âge comparable. Par contre les résultats qui se rapportent à la compréhension sont beaucoup plus incertains. L'auteur écrira :

"The writer's personal belief is that the differences in silent reading comprehension between German and American children are unstable and probably reversible" (Preston, 1963).

Quant à l'incidence des échecs dans l'apprentissage de la lecture, elle semble équivalente dans les deux pays (Bond et Tinker, 1957 ; Preston, 1963 ; Samuels, 1969).

Dailey (1975) qui a comparé le rendement en lecture d'étudiants de la troisième à la sixième année dont une partie avait fait son apprentissage de la lecture à partir d'un alphabet phonétique et l'autre à partir de l'alphabet anglais traditionnel conclut qu'il n'a trouvé aucune différence significative dans le rendement en lecture chez les enfants des différents niveaux pris en considération.

Une étude menée par Woodworth (1938) auprès d'adolescents chinois et américains a montré que dans le cas d'une langue idéographique et dans celui d'une langue à code phono-graphique, le rendement en lecture, quand il est calculé en terme de contenu lu par unité de temps, est le même.

Bref, l'étude de Gibson et al. suggérait que les apprentis-lecteurs assimileraient des règles de correspondance entre des suites graphémiques et des suites phonologiques de plus

en plus complexes. Cette hypothèse implique que les langues dans lesquelles les correspondances suites graphémiques-suites phonologiques sont simples permettent un apprentissage plus rapide de la lecture. Or de nombreuses études comparant les codes phono-graphiques simples à des codes complexes démontrent en effet que les langues à code phono-graphique simple permettent un apprentissage plus rapide de la lecture, ce qui semblerait confirmer l'hypothèse de Gibson et al.. Cependant, lorsque les apprentis-lecteurs des langues à code phono-graphique complexe ont eu suffisamment de temps pour maîtriser toutes les correspondances, ils peuvent lire aussi efficacement que les autres.

1.7.2 Vérification de l'hypothèse L2 chez les apprentis-lecteurs

D'autres études donnent des indications sur la façon dont les enfants acquièrent une maîtrise progressive de la grammaire des suites graphémiques correctes dans une langue donnée.

Déjà en 1927, Hoffman remarquait que des groupes de consonnes dont l'agencement ne se rapprochait pas d'un agencement possible en anglais n'étaient pas lus plus rapidement par des adultes que par des enfants. Par contre plus ces groupes de consonnes étaient près de configurations anglaises, plus l'avantage des adultes sur les enfants devenait

évident.

A la suite d'une étude qui voulait vérifier les habitudes d'association de lettres chez les adultes et les enfants, Amster et Keppel concluent:

> "Frequencies of specific letter responses for adults tended on the whole to conform more closely to the frequencies in the language than did the children's frequencies" (Amster, Keppel, 1968, p. 330).

Alors que les enfants de première année à qui on demande de classer des mots fictifs dans une boîte de "vrais mots" et de "faux mots" le font au hasard, les enfants de troisième jugeraient correctement de la ressemblance des mots dans une proportion de 80% (Rosinski, Wheeler, 1972). En troisième année, les enfants auraient donc déjà suffisamment intégré la grammaire de l'orthographe lexicale pour distinguer entre les configurations possibles et celles qui ne le sont pas alors qu'en première année, cette habileté n'existerait pas encore.

Une autre étude, celle de Lefton, Spragins et Byrnes (1973), illustre aussi cette évolution du contrôle de la grammaire de l'orthographe en comparant les performances d'enfants encore au début de leur scolarité et celles d'autres enfants ayant une scolarité plus longue. Leur expérience consistait à présenter à des écoliers de première, troisième et cinquième années, des groupes de mots fictifs dont

certains ressemblaient à des mots anglais et d'autres moins. Une lettre manquait dans chacun de ces mots et les enfants devaient trouver laquelle. Les résultats ont démontré que plus l'expérience en lecture est longue, plus les contraintes et les règles de la langue écrite sont utilisées pour solutionner le problème posé. Alors que les enfants de première année donnaient des réponses un peu au hasard, on pouvait se rendre compte que ceux de troisième année connaissaient des indices qui leur permettaient de fournir des réponses sensées. Ceux de cinquième année ont obtenu les meilleurs résultats.

A la suite de ses travaux portant sur l'apprentissage de la lecture, Santa(1976) a obtenu des résultats comparables. Elle écrit:

> "The development of reading maturity appears to correspond with more flexible word recognition strategies and an increased dependence on larger units of analysis" (Santa, 1976)

Afin de savoir comment se fait l'intégration des règles de la langue écrite, Gibson entreprend, avec la collaboration de Amidon (1970) une nouvelle étude dans laquelle elle compare les performances des enfants dans la solution de problèmes de décodage de mots fictifs avec celles qu'ils obtiennent dans des tâches analogues mais dans lesquelles les lettres sont remplacées par des bâtonnets de différentes cou-

leurs. Elle observe d'abord que les apprentis-lecteurs semblent clairement développer un mécanisme de recherche de configurations régulières ou constantes et que cette habileté se développerait avec l'âge et l'habitude de lire. De plus, les résultats obtenus indiquent que ce mécanisme de recherche peut être stimulé par une approche adéquate. Une fois acquis, il pourrait être utilisé pour résoudre de nouveaux problèmes.

Des études qui précèdent, on peut donc déduire qu'il y aurait, au cours de l'apprentissage de la lecture, une découverte progressive des règles de la langue écrite qui déterminent l'acceptabilité de suites graphémiques. L'hypothèse L2 est ainsi confirmée chez les apprentis-lecteurs.

Au cours de son apprentissage de la lecture, l'enfant apprendrait donc à maîtriser au moins deux habiletés. D'une part, il apprendrait à associer une suite graphémique à la suite phonologique correspondante (Hypothèse L1). D'autre part, il saurait progressivement reconnaître les suites graphémiques correctes dans la langue qu'il apprend à lire (Hypothèse L2). Qu'en est-il de l'apprentissage de l'orthographe et plus particulièrement de l'orthographe lexicale? Les études qui ont porté sur l'apprentissage de l'orthographe ne renseignent pas sur la façon dont se fait cet apprentissage. Pour formuler des hypothèses de travail, on a eu recours aux conclusions des études portant sur l'apprentissa-

ge de la lecture en espérant qu'elles pouvaient constituer un point de départ valable malgré les différences que présentent les habiletés à lire et à orthographier.

1.8 L'apprentissage de l'orthographe lexicale: hypothèses.

La connaissance du code phono-graphique qui a été défini par Blanche-Benveniste et Chervel (1969), comme "l'ensemble des règles de correspondance entre graphèmes et phonèmes pratiquées par un individu, un milieu ou une époque", contribuerait à la maîtrise de l'orthographe lexicale. Cette définition que donnent Blanche-Benveniste et Chervel du code phono-graphique est toutefois ambigüe puisqu'elle ne précise pas tous les éléments qui interviennent dans ces règles. En effet, les règles de correspondance phono-graphique du français ne font pas qu'associer un graphème à un phonème. Elles déterminent quels graphèmes peuvent être utilisés pour représenter un phonème en tenant compte de l'environnement phonologique de ce phonème et de l'environnement graphémique du graphème qui doit le représenter. On pourrait donc redéfinir de façon plus précise le code phono-graphique comme "l'ensemble des règles de correspondance entre des suites graphémiques et des suites phonologiques".

Il ne semble pas qu'il ait été fait, jusqu'à maintenant, des études sur l'acquisition du code phono-graphique. Il serait cependant logique de penser que les personnes qui apprennent à orthographier le français apprendraient d'abord à maîtriser les relations les plus simples, c'est-à-dire les relations qui impliquent un nombre minimum d'éléments comme,

par exemple, un phonème et les graphèmes qui peuvent le représenter. Il serait aussi logique de croire que peu à peu, l'enfant maîtriserait des relations plus complexes. Ainsi, il établirait des correspondances entre des suites graphémiques et des suites phonologiques de plus en plus longues et de moins en moins fréquentes. Parallèlement, l'enfant apprendrait aussi à distinguer les suites graphémiques qui peuvent correctement représenter une suite phonologique en français.

Les études sur la lecture semblent démontrer que le lecteur aurait recours à un processus similaire. En effet, les études sur la lecture indiquent que l'exercice de cette habileté impliquerait le recours à au moins deux types de connaissances. D'une part, le lecteur doit savoir associer une suite graphémique donnée à la suite phonologique correspondante. D'autre part, il apprend aussi à découper les mots de façon à identifier les suites graphémiques caractéristiques du français. La relation entre ces deux compétences n'est pas clairement établie. On pourrait cependant penser que la seconde, c'est-à-dire l'identification des suites graphémiques bien formées, permettrait de réaliser la première qui consiste à associer ces suites graphémiques aux suites phonologiques appropriées.

Il paraît plausible que la maîtrise de ces habiletés ait un impact sur la compétence en écriture des individus.

Par exemple, le fait qu'un individu puisse bien identifier les suites graphémiques bien formées dans un texte devrait l'aider à éviter de produire des suites mal formées lorsqu'il écrit. De la même façon un individu qui sait bien associer une suite phonologique à une suite graphémique donnée devrait pouvoir bénéficier d'une façon ou d'une autre de cette compétence lorsque vient le moment d'associer une suite graphémique à une suite phonologique.

A cause de l'absence quasi-totale d'études sur les mécanismes d'acquisition de l'orthographe, on ne sait pas comment se fait l'apprentissage de l'habileté à orthographier. Il est probable que la mémoire visuelle des mots intervienne d'une certaine façon dans l'apprentissage de l'orthographe. Toutefois, à la lumière des études sur l'apprentissage de la lecture et des constatations sur les relations entre l'écriture et la lecture, il ne paraît pas invraisemblable de penser que celui qui apprend à orthographier dans une langue suivrait un processus semblable à celui qui apprend à lire. On est ainsi amené à énoncer les trois hypothèses de travail E1, E2 et E3 qui s'apparentent aux hypothèses L1 et L2 au sujet de la lecture.

Puisque l'apprenti-lecteur n'apprend pas à lire seulement en mémorisant la configuration lexicale de chaque mot, on peut supposer que, corollairement, celui qui apprend à

orthographier le français ne ferait pas cet apprentissage par la seule mémorisation visuelle des mots. Pour s'en assurer, on a voulu tester l'hypothèse suivante:

HYPOTHÈSE E1 : L'apprentissage de l'orthographe lexicale ne se fait pas seulement par la mémorisation de la représentation graphémique de chacun des mots de la langue.

En français, la plupart des phonèmes peuvent être représentés par plus d'une expression graphémique. Certains peuvent même l'être par plus de vingt graphies différentes. Pour qu'une personne parvienne à écrire correctement en français, on peut penser qu'elle acquiert progressivement la maîtrise de ces associations. L'hypothèse E2 explicite ce phénomène:

HYPOTHÈSE E2 : L'individu qui fait l'apprentissage de l'orthographe lexicale française apprend progressivement à associer, à chaque phonème, les diverses expressions graphémiques qui peuvent le représenter.

Il existe, en français, tout un ensemble de règles qui déterminent lesquelles, parmi les expressions graphémiques qui représentent un phonème, peuvent être utilisées, compte tenu des autres graphèmes qui les entourent. Cette caractéristique de la langue française, de même que les conclusions des études sur l'apprentissage de la lecture, suggèrent l'hypothèse suivante sur l'apprentissage de l'orthographe:

HYPOTHESE E3 : L'individu qui apprend à orthographier le français acquiert, au moins inconsciemment, la maîtrise progressive des règles de formation des suites graphémiques correctes dans cette langue.

Chacune des expressions graphémiques qui peut représenter un phonème n'apparaît pas, en français, avec la même fréquence. Certaines sont en effet beaucoup plus fréquentes que d'autres. L'expression graphémique ss, par exemple, représente le phonème /s/ en position médiane dans environ 30% des cas. La graphie ç n'est utilisée par contre, pour représenter ce phonème, que dans environ 5% des cas. Au cours de son apprentissage, un individu qui apprend à orthographier le français intégrerait peut-être, d'une façon plus ou moins consciente, la fréquence des expressions graphémiques. Devant le problème de la représentation d'un phonème auquel correspondent plusieurs graphèmes, on peut conjecturer qu'il utiliserait le plus souvent les expressions graphémiques qui représentent le plus fréquemment ce phonème. On est ainsi amené à formuler une quatrième hypothèse qui se lit comme suit:

HYPOTHESE E4 : L'individu qui apprend à orthographier le français sur-utilise, pour représenter un phonème, les expressions graphémiques les plus fréquentes de ce phonème.

Pour vérifier l'hypothèse E4, on ne devrait retenir que les individus qui ne connaîtraient pas les règles de formation des suites graphémiques, c'est-à-dire, les individus qui auraient un score faible lors de la vérification de l'hypothèse E3. En effet, les individus qui maîtrisent les règles de formation des suites graphémiques sauraient quelle graphie utiliser pour représenter un phonème dans un contexte donné et reproduiraient conséquemment les fréquences théoriques des graphies du français.

La confirmation des hypothèses E1, E2, E3 et E4 impliquerait que l'apprentissage de l'orthographe lexicale ne se ferait clairement pas par la seule mémorisation de l'orthographe de chacun des mots de la langue. Dans ce cas, les vocabulaires de base et les échelles de difficulté qui ont découlé des recherches pédagogiques sur l'orthographe deviendraient des outils dont la pertinence, pour l'apprentissage de l'orthographe lexicale, pourrait être discutée.

Pour vérifier les quatre hypothèses de travail (E1, E2, E3, E4), on a effectué une étude expérimentale dont la méthodologie et les résultats sont présentés dans les chapitres subséquents.

CHAPITRE II

MÉTHODOLOGIE

2.1 Introduction

Le code phono-graphique a été défini dans cette étude comme l'ensemble des règles de correspondance entre suites graphémiques et suites phonologiques. Il découle de cette définition que plus le code phono-graphique d'un individu se rapproche du code phono-graphique français, meilleure devrait être la maîtrise de l'orthographe lexicale française de cet individu.

L'hypothèse E1 pose que la maîtrise de l'orthographe lexicale ne se fait pas uniquement par la mémorisation de la représentation graphémique de chacun des mots de la langue. Les hypothèses E2 et E3 décrivent des habiletés qui interviendraient dans l'apprentissage de l'orthographe lexicale. Elles suggèrent qu'une personne qui apprend à orthographier le français acquiert progressivement, d'une part, la maîtrise des représentations graphémiques des phonèmes de cette langue et, d'autre part, celle des règles de formation des suites graphémiques correctes dans cette langue. Quant à l'hypothèse E4, elle propose que l'individu utili-

se, pour représenter un phonème, des expressions graphémiques dont la fréquence est en corrélation avec la fréquence de ces expressions graphémiques dans la langue française. C'est en faisant l'acquisition des habiletés ou des stratégies décrites par les hypothèses E2, E3 et E4 qu'un individu parviendrait à assimiler le code phono-graphique français.

Pour vérifier les quatre hypothèses de travail, il a semblé pertinent d'obtenir une production écrite d'individus qui apprennent à orthographier le français. Cette production sera obtenue au moyen d'un test de mots fictifs dont la construction est présentée dans la section qui suit. Les caractéristiques des sujets de l'échantillon retenus pour les fins de cette étude sont présentées dans la section 2.3. La dernière partie du chapitre explique la procédure suivie pour obtenir les données qui serviront à vérifier les hypothèses de travail.

2.2 Construction de l'épreuve orthographique

2.2.1 Contraintes imposées par les hypothèses

Pour vérifier les quatre hypothèses de travail, il a semblé approprié d'obtenir une production de mots écrits d'individus qui apprennent à orthographier le français. On aurait pu avoir recours à une production libre, chaque sujet rédigeant un texte sur un thème donné par exemple. On aurait pu aussi demander aux élèves de compléter des phrases avec des mots de leur choix ou encore utiliser tout autre moyen qui aurait permis d'obtenir un corpus de mots qui auraient pu varier d'un étudiant à l'autre. Cependant, pour vérifier de façon plus rigoureuse les hypothèses de travail, il a semblé préférable d'obtenir un corpus fermé. Aussi a-t-on construit une épreuve orthographique à laquelle on a soumis de façon uniforme tous les sujets de l'échantillon. La construction de l'épreuve a nécessité que l'on tienne compte de certaines contraintes imposées par la nature des hypothèses à vérifier et par les outils mathématiques qui seront utilisés lors de la vérification de ces hypothèses. Le reste de cette section montre comment les hypothèses de travail ont déterminé la construction de l'épreuve.

L'hypothèse E1 se lit comme suit:

HYPOTHÈSE E1: L'apprentissage de l'orthographe lexicale ne se fait pas seulement par la mémorisation de la représentation graphémique de chacun des mots de la langue.

Pour savoir si le code phono-graphique des sujets soumis à l'épreuve est le produit de la mémorisation des mots de la langue (hypothèse E1) ou celui des habiletés ou stratégies décrites par les hypothèses E2, E3 et E4, on a construit un test de mots fictifs ou logatomes.(1) Le recours à des logatomes présente en effet l'avantage de neutraliser l'influence de la mémorisation des mots.

Le code phono-graphique français est constitué d'un ensemble de règles de correspondance entre suites graphémiques et suites phonologiques. Ces règles font que dans un texte français, chacune des différentes graphies utilisées pour représenter un phonème apparaît avec une fréquence qui, exprimée en pourcentage, tend vers une valeur donnée que l'on nomme, dans cette étude, fréquence théorique en français. Celui qui connaît certaines règles du code phono-graphique français les applique lorsqu'on lui demande d'écrire des mots français ou des mots présentés comme tels, comme c'est le cas avec les logatomes de l'épreuve orthographique. Conséquemment, il utilise les diverses graphies avec des fréquences qui tendent vers leur fréquence théorique en français. Par ailleurs, celui qui ne connaît pas ces règles ne respectera vraisemblablement pas les fréquences théoriques des expressions graphémiques.

---

(1) "On appelle logatome une syllabe ou une suite de syllabes appartenant à une langue mais ne formant pas un mot ou un syntagme significatif" (Dubois et al. 1973, p. 305).

Pour vérifier s'il se fait un apprentissage qui n'est pas dû à la mémorisation, on aura recours à une première mesure, l'indice de similitude, qui consiste essentiellement à comparer les fréquences des graphies utilisées par un individu avec leur fréquence théorique en français. L'apparition des expressions graphémiques françaises étant dictée par leur environnement phonologique et graphémique, on a dû construire le test de façon à respecter les fréquences des suites phonologiques des phonèmes qui seront étudiés.

L'hypothèse E2 se lit comme suit:

HYPOTHÈSE E2 : L'individu qui fait l'apprentissage de l'orthographe lexicale française apprend progressivement à associer, à chaque phonème, les diverses expressions graphémiques qui peuvent le représenter.

Pour vérifier cette hypothèse, on aurait pu demander à chacun des sujets de l'échantillon d'écrire toutes les expressions graphémiques qu'il connaît pour représenter un phonème. Il est cependant probable que plusieurs élèves auraient oublié de reproduire, avec cette méthode, des expressions graphémiques pertinentes qu'ils connaissent. On aurait aussi pu demander aux étudiants d'indiquer, parmi une liste d'expressions graphémiques, lesquelles peuvent représenter un phonème donné. L'inconvénient de cette fa-

çon de procéder est de vérifier une habileté davantage reliée à la lecture qu'à l'écriture. Il a finalement été jugé préférable de présenter, aux étudiants, l'occasion d'utiliser les graphies qu'ils connaissent en leur donnant la dictée de logatomes dans lesquels les phonèmes étudiés apparaissent dans des environnements qui permettent l'utilisation de diverses expressions graphémiques.

Or il existe, en français, des phonèmes qui ne possèdent que quelques expressions graphémiques. C'est le cas par exemple de /v/ qui peut être représenté par v, ve et w. Certains phonèmes, par contre, possèdent de nombreuses expressions graphémiques. Le phonème /ɛ/ par exemple en possède au moins vingt-quatre, le phonème /e/, treize, le phonème /ã/, dix-huit, le phonème /ɛ̃/, vingt-trois, etc. (Burney, 1959, p. 39). Pour vérifier l'hypothèse E2, il a paru préférable d'avoir recours à des phonèmes auxquels correspondent de nombreuses expressions graphémiques. Il sera ainsi plus facile de savoir comment évolue la maîtrise des représentations graphémiques chez celui qui apprend à orthographier le français.

Certaines représentations graphémiques des phonèmes qui en possèdent un grand nombre n'apparaissent que dans certaines positions à l'intérieur d'un mot. Par exemple le phonème /ɛ/ n'est jamais représenté par la suite graphémique *ei* en début d'un mot. Par ailleurs, /s/ ne s'écrit

jamais ç s'il n'est pas suivi des graphèmes a, o ou u. Ainsi, afin de placer l'individu qui apprend à orthographier le français dans une situation qui favorise le recours aux diverses représentations graphémiques d'un phonème, l'environnement de ce phonème devra être contrôlé.

Ce contrôle de l'environnement des phonèmes permettra de vérifier également l'hypothèse E3 qui se lit comme suit :

HYPOTHÈSE E3 : L'individu qui apprend à orthographier le français acquiert au moins inconsciemment la maîtrise progressive des règles de formation des suites graphémiques correctes dans cette langue.

Il existe en effet, en français, tout un ensemble de règles qui déterminent laquelle, parmi les expressions graphémiques qui représentent un phonème, doit être utilisée, compte tenu des autres graphèmes qui l'entourent. Pour savoir si l'enfant a acquis la maîtrise de ces règles, c'est-à-dire pour déterminer si les expressions graphémiques qu'il aura choisies pour représenter des phonèmes sont correctes, il sera nécessaire de tenir compte de leurs environnements. Par ailleurs, les environnements devront avoir été sélectionnés de façon à ce que des règles de formation des suites graphémiques correctes puissent être appliquées.

La vérification de l'hypothèse E4 n'impose, quant à elle, aucune contrainte particulière dans la construction de l'épreuve orthographique. Cette hypothèse se lit comme suit :

HYPOTHÈSE E4: L'individu qui apprend à orthographier le français sur-utilise, pour représenter un phonème, les expressions graphémiques les plus fréquentes de ce phonème.

Pour que les données recueillies par l'épreuve orthographique puissent être traitées statistiquement, chacun des phonèmes retenus devra apparaître suffisamment souvent dans l'épreuve orthographique. Par ailleurs, celle-ci devra être d'une longueur qui soit adaptée au jeune âge de certains des sujets de l'échantillon. Pour satisfaire à ces deux conditions, le nombre de phonèmes étudiés devra donc être restreint.

On a retenu les phonèmes /s/, /ɛ/ et /o/. Ce choix est relativement arbitraire. Il répond toutefois aux exigences imposées par l'hypothèse E2 puisqu'à chacun de ces phonèmes correspondent de nombreuses représentations graphémiques. Il répond aussi à celles de l'hypothèse E3 puisque des règles régissent l'utilisation des expressions

graphémiques de ces phonèmes.

Avant de mettre au point l'épreuve orthographique, il reste à tenir compte d'une dernière contrainte imposée par l'hypothèse E1 et qui consiste à établir la fréquence, dans la langue française, des expressions graphémiques des phonèmes étudiés.

2.2.2 Fréquence des expressions graphémiques des phonèmes /s/, /ɛ/ et /o/ en français

La vérification de l'hypothèse E1 implique que soit connue la fréquence, dans la langue française, des expressions graphémiques des phonèmes étudiés. Cette information est également nécessaire pour la vérification de l'hypothèse E4 qui veut établir le rapport entre les fréquences des expressions graphémiques utilisées par les individus et leurs fréquences théoriques en français.

Puisque l'on s'apprête à comparer la fréquence des expressions graphémiques utilisées par les enfants dans l'épreuve orthographique à leur fréquence dans la langue française, il importe de calculer ces fréquences théoriques à partir du sous-ensemble du français connu par les enfants.

Le dépouillement de tout le matériel auquel les enfants sont exposés - manuels de lecture, livres d'histoires, etc. - aurait été fort fastidieux et peut-être inutile. Il existe, en effet, des lexiques du vocabulaire des enfants du primaire. La principale faiblesse de ces lexiques, compte tenu des besoins de cette étude, est qu'ils ne rapportent pas la fréquence des mots lus par les enfants mais celle des mots utilisés dans le langage oral et écrit. Or les livres de lecture pour enfants abordent le plus souvent des sujets connus des enfants en ayant ordinairement recours à des mots qu'ils connaissent. Quand le vocabulaire qui y est utilisé ne leur est pas déjà familier, le livre devient alors une occasion pour les enfants de l'apprendre. Les mots de ce nouveau vocabulaire sont donc, par la suite, susceptibles de faire partie de leur vocabulaire actif. On peut donc supposer que les mots utilisés par les enfants et les mots lus constituent des ensembles qui tendent à se recouvrir, surtout dans le cas des mots les plus fréquents. Les lexiques existants du vocabulaire des

élèves du primaire peuvent donc assez bien représenter le sous-ensemble de la langue française connu des enfants.

On a retenu, parmi les lexiques existants, celui de Préfontaine et al. (1968). Il constitue un inventaire du vocabulaire oral des élèves de la maternelle à la troisième année. Pour construire ce lexique, le vocabulaire des enfants a été recueilli durant une année scolaire entière, ce qui a permis de capter de nombreux centres d'intérêt de l'enfant : saisons, sports, habillement, nourriture, vie quotidienne, voyages, vie familiale, etc. De plus, le grand nombre de classes dans lesquelles le vocabulaire a été recueilli, les diverses régions géographiques ainsi que l'éventail de milieux culturels et économiques auxquels ces classes d'étudiants appartiennent permettent de croire que l'échantillon retenu était assez représentatif de la jeune population scolaire du Québec.

Le principal problème que le choix de ce lexique pose est celui de ne contenir que le vocabulaire oral des enfants du premier cycle du primaire. Il ressort toutefois du dépouillement d'un second lexique, celui de Doucet (1971), dans lequel a été recueilli le vocabulaire écrit des élèves de douze classes de cinquième année, que la fréquence des expressions graphémiques des phonèmes étudiés n'est pas significativement différente de celles qui pro-

viennent du dépouillement du lexique de Préfontaine et al. C'est donc à partir de ce lexique que la fréquence des expressions graphémiques des phonèmes /s/, /ɛ/ et /o/ en français sera établie.

Certaines expressions graphémiques ne peuvent apparaître, dans un mot, qu'en position initiale alors que d'autres n'apparaissent qu'en position médiane ou finale. La position initiale a été définie comme étant celle qui est occupée par l'expression graphémique qui représente le premier phonème d'un mot. Dans le mot *manteau* par exemple, *m* est en position initiale. La position médiane a été définie comme étant celle qui est occupée par l'expression graphémique qui représente un phonème qui n'est ni le premier, ni le dernier d'un mot. Dans le mot *manteau*, *an* et *t* sont en position médiane. La position finale a été définie comme étant celle qui est occupée par l'expression graphémique qui représente le dernier phonème d'un mot. Dans le mot *manteau*, *eau* est en position finale.

A cause de cette variation de la fréquence des expressions graphémiques selon la position dans un mot, le dépouillement du lexique a été effectué en relevant chacune des expressions graphémiques représentant l'un ou l'autre des phonèmes retenus, en la classant selon la position qu'elle occupe dans le mot et en y faisant correspondre la fréquence du mot duquel elle fait partie. On a ain-

si obtenu la fréquence des expressions graphémiques des phonèmes étudiés. Les résultats de ce dépouillement sont présentés aux tableaux 2.01, 2.02 et 2.03.

2.2.3 L'épreuve orthographique

Il ressort, de l'examen des tableaux 2.01, 2.02 et 2.03, que la fréquence des expressions graphémiques varie en fonction de leur position dans le mot. ç par exemple n'apparaît qu'en position médiane, *os* qu'en position finale. Par ailleurs, ces tableaux de fréquences doivent servir entre autres à déterminer si la fréquence des expressions graphémiques utilisées par un individu pour représenter un phonème se rapproche de la fréquence de ces expressions graphémiques dans la langue française.

A cause de la nature des données et de l'objectif poursuivi, le test du khi deux semble tout indiqué pour fournir cette information. En comparant les fréquences théoriques aux fréquences empiriques, ce test détermine s'il existe une différence statistiquement significative entre les deux ensembles de fréquences. Or pour qu'il soit valable, ce test exige une fréquence théorique minimum que plusieurs spécialistes (Dayhaw, 1969, p. 379; Bertaud, 1971, p. 161) établissent à cinq.

L'examen des tableaux 2.01, 2.02 et 2.03 indique que la répartition des fréquences théoriques correspon-

TABLEAU 2.01

FRÉQUENCES EN FRANCAIS DES EXPRESSIONS GRAPHEMIQUES DU PHONÈME /s/ EN FONCTION DE SA POSITION DANS UN MOT

EXPRESSIONS GRAPHÉMIQUES	POSITIONS		
	INITIALE %	MÉDIANE %	FINALE %
s	86,5	32,4	19,2
c	12,6	22,7	32,7
ss	0,0	32,0	43,0
ç	0,0	6,5	0,0
autres	0,9	6,4	5,1
TOTAL	100,0	100,0	100,0

TABLEAU 2.02

FRÉQUENCES EN FRANCAIS DES EXPRESSIONS GRAPHÉMIQUES DU PHONÈME /ɛ/ EN FONCTION DE SA POSITION DANS UN MOT

EXPRESSIONS GRAPHÉMIQUES	POSITIONS		
	INITIALE %	MÉDIANE %	FINALE %
è	50,8	6,5	9,4
e	18,5	51,3	0,0
ai	27,5	20,8	0,0
ê	0,0	13,1	0,0
êc	0,0	0,0	13,3
et	0,0	0,0	43,5
ais	0,0	0,0	10,3
ait	0,0	0,0	9,0
autres	3,2	8,3	14,5
TOTAL	100,0	100,0	100,0

TABLEAU 2.03

FRÉQUENCES EN FRANCAIS DES EXPRESSIONS GRAPHEMIQUES DU PHONÈME /o/ EN FONCTION DE SA POSITION DANS UN MOT

EXPRESSIONS GRAPHÉMIQUES	POSITIONS		
	INITIALE %	MÉDIANE %	FINALE %
au	89,5	36,7	0,0
os	0,0	0,0	7,5
o	0,0	44,9	12,5
eau	0,0	9,3	61,9
ô	0,0	8,9	0,0
ot	0,0	0,0	7,3
autres	10,5	2,0	11,1
TOTAL	100,0	100,0	100,0

dant aux diverses expressions graphémiques des phonèmes retenus est très différente d'une position à l'autre, étant parfois très faible ou même égale à zéro dans plusieurs cas. Une représentation satisfaisante des expressions graphémiques de chacune des positions dans le mot aboutirait à la construction d'une épreuve orthographique démesurément longue.

Par contre, en éliminant les expressions graphémiques des phonèmes en position médiane auxquelles correspond la valeur zéro, il reste encore dans cette position un minimum de quatre expressions graphémiques pour chacun des phonèmes retenus. Ce nombre est suffisant pour vérifier l'hypothèse E2. De plus, la fréquence la plus faible est de 6.5%. En fixant à cinq la fréquence théorique brute minimum de chacune des expressions graphémiques des phonèmes retenus en position médiane, il faudrait construire une épreuve orthographique dans laquelle les phonèmes /s/ et /ɛ/ apparaîtraient soixante-dix-sept fois et le phonème /o/, cinquante-six fois. Ce nombre d'occurrences paraît approprié pour construire une épreuve orthographique qui ne soit pas d'une longueur telle qu'elle déborde les contraintes imposées par les administrations scolaires et le jeune âge de certains des sujets de l'échantillon. Ce sont donc les fréquences des expressions graphémiques des phonèmes /s/, /ɛ/ et /o/ en position médiane qui serviront de base à la construction de l'épreuve orthographique.

Celle-ci comportera soixante-dix-sept occurrences de chacun des phonèmes /s/ et /ɛ/ et cinquante-six du phonème /o/.

Le choix d'une expression graphémique pour représenter une phonème peut varier en fonction de l'environnement de ce phonème. Il a déjà été établi que pour vérifier les hypothèses E2 et E3, il faudra tenir compte des environnements des phonèmes retenus dans la construction de l'épreuve orthographique. Par ailleurs, la vérification de l'hypothèse E1 exige que les environnements des phonèmes étudiés apparaissent dans des proportions qui tiennent compte de leur fréquence dans le sous-ensemble de la langue française utilisé par les enfants.

Or le nombre d'environnements différents pour chacun des phonèmes retenus est plus grand que le nombre d'occurrences du phonème dans l'épreuve orthographique. Il a donc fallu laisser de côté les environnements les moins fréquents. Ainsi, puisque le phonème /s/ compte soixante-dix-sept occurrences dans la dictée, une occurrence équivaut à .013 des occurrences de /s/. On a donc conservé les environnements les plus fréquents de sorte que les moins fréquents parmi les environnements retenus apparaissent au moins une fois dans l'épreuve orthographique. Puisqu'une occurrence du phonème /s/ équivaut à .013 du nombre d'occurrences de ce phonème, on a conservé les environnements

dont la proportion d'apparition dans le corpus est égale ou supérieure à .013 de l'ensemble des occurrences des environnements retenus. C'est ainsi qu'on n'a retenu que les vingt-neuf environnements les plus fréquents du phonème /s/. Le moins fréquent parmi ceux-ci a, dans le corpus de Préfontaine et al., une fréquence de quarante-neuf, ce qui correspond à .013 de l'ensemble des vingt-neuf environnements les plus fréquents du phonème /s/ qui totalisent, dans le corpus retenu, 3743 occurrences de ce phonème en position médiane.
(49 ÷ 3743 = .013)

Pour établir le nombre d'occurrences de chacun des environnements affectant l'un des trois phonèmes retenus, on a multiplié la proportion de cet environnement par rapport à l'ensemble des environnements retenus pour un des phonèmes par le nombre d'occurrences de ce phonème dans l'épreuve orthographique.

La liste de environnements phonologiques conservés pour chacun des phonèmes ainsi que le nombre d'occurrences de chacun des environnements sont présentés aux tableaux 2.04, 2.05 et 2.06.

Pour que l'épreuve orthographique soit la plus courte possible, le moyen le plus économique pour construire les

logatomes consistait à accoler les triphones présentés aux tableaux 2.04, 2.05 et 2.06 de façon à constituer des mots d'au plus trois syllabes et dont la consonnance se rapproche le plus possible de celle de mots français. L'épreuve orthographique finalement obtenue est composée de cent trois logatomes. Elle est présentée au tableau 2.07.

TABLEAU 2.04

NOMBRE D'OCCURRENCES DE CHACUN DES ENVIRONNEMENTS PHONOLOGIQUES DU PHONÈME /s/ CONSERVÉS POUR LA CONSTRUCTION DE L'ÉPREUVE ORTHOGRAPHIQUE

ENVIRONNEMENTS PHONOLOGIQUES	NOMBRE D'OCCURRENCES	ENVIRONNEMENTS PHONOLOGIQUES	NOMBRE D'OCCURRENCES
ase	12	ɛsɛ	2
ãse	8	asw	2
œsj	6	use	2
rsɔ̃	5	ɛ̃st	1
lsk	4	ɛsp	1
asɔ̃	4	rso	1
lsi	4	rsi	1
lsj	3	asœ	1
ase	2	ɛsœ	1
esi	2	ysɔ̃	1
rse	2	usɛ̃	1
ɛst	2	ɔ̃st	1
asi	2	lso	1
rsj	2	ysɔ̃	1
rsɛ	2		

TABLEAU 2.05

NOMBRE D'OCCURRENCES DE CHACUN DES ENVIRONNEMENTS PHONOLOGIQUES DU PHONÈME /ɛ/ CONSERVÉS POUR LA CONSTRUCTION DE L'ÉPREUVE ORTHOGRAPHIQUE

ENVIRONNEMENTS PHONOLOGIQUES	NOMBRE D'OCCURRENCES	ENVIRONNEMENTS PHONOLOGIQUES	NOMBRE D'OCCURRENCES
fɛr	13	vɛj	2
mɛr	5	fɛt	2
pɛt	5	pɛʃ	2
mɛt	5	nɛt	2
ʃɛr	4	rɛn	2
tɛr	4	dɛl	2
vɛr	4	jɛl	1
mɛz	4	tɛj	1
ʃɛr	3	bɛn	1
vɛk	3	sɛl	1
rɛj	2	dɛs	1
rɛr	2	rɛs	1
nɛz	2	rɛz	1
lɛj	2		

TABLEAU 2.06

NOMBRE D'OCCURRENCES DE CHACUN DES ENVIRONNEMENTS PHONOLOGIQUES DU PHONÈME /o/ CONSERVÉS POUR LA CONSTRUCTION DE L'ÉPREUVE ORTHOGRAPHIQUE

ENVIRONNEMENTS PHONOLOGIQUES	NOMBRE D'OCCURRENCES	ENVIRONNEMENTS PHONOLOGIQUES	NOMBRE D'OCCURRENCES
sot	9	pov	2
bok	7	kom	2
sov	4	lok	2
fot	4	mot	1
ʒon	3	ʃos	1
tom	3	rol	1
ros	3	kok	1
poz	3	kod	1
kot	2	rot	1
tob	2	rok	1
ʃod	2	fog	1

TABLEAU 2.07

L'ÉPREUVE ORTHOGRAPHIQUE

mersə	rokusɛ̃	tɛrsə	komisk
rosfɛr	kodasə	roluse	ʃɛrsɔ̃
beŋãse	fotysi	komusə	ferasə
kokasɔ̃	tomœsjã	pɛrɛn	pozvɛj
fɛrdɛl	metrœsjã	lokɛsə	fɛrsə
bokəsi	fɛrmɛt	ʃɛrɛj	fogɛsɛ
fɛrɛn	verymɛz	pɛrœsjã	ʃosjɛr
mɛtobr	lskesɛ	sotœsjɔ̃	asəpɛr
rɛsəsə	sovɛk	mɛrsalã	tomyʃɛr
rɛzlso	pozœsjo	pɛrfod	bokasɔ̃
sotəfɛt	ɔ̃stɛr	fotafɛt	rotasə
bokãsə	sotãse	vɛkot	mɛzœsjã
sotasik	rɛrasə	sovɛj	pɛrsɔ̃t
dɛlɛj	pɛrfod	rozlsij	sotasə
pozlsil	fotlsin	avɛrsɔ̃	asisisk
ʒonmɛr	rɛjɛst	tɛrɛr	fɛrsi
boklsjɔ̃	fɛrlɛj	nɛʒoni	fɛtasə
mɛtrosɔ̃	asəsik	ɛstob	fɛrãsə
nɛʒon	pɛʃáse	boklsk	sovɛrã
sotasɔ̃	sotom	jɛlysɔ̃	bokasə
dɛsot	ʃodasə	rəsənɛt	povɛk
sɛlasə	fɛrsjɛr	sotlnɛt	mɛrsjɔ̃
povɛr	tɛjasœr	rəsəmɛr	fɛrsɔ̃
aswɛsœr	tɛrsɔ̃	mɛtãsə	fɛrso
bokɛ̃stə	lsljɛr	sovãsə	mɛzãsə
lokisjɛr	motasɔ̃	mɛzkotɛ̃	

2.3 Caractéristiques de l'échantillon

Pour vérifier les hypothèses, l'épreuve orthographique a été soumise à un échantillon de cent quatre-vingt-quinze écoliers du primaire dont trente-cinq de deuxième année et quarante de chacun des niveaux de la troisième à la sixième année. Cet échantillon provient de deux écoles dans chacune desquelles les classes sont constituées d'élèves forts, moyens et faibles. Pour faire partie de l'échantillon, les sujets devaient être de langue maternelle française. Ils devaient de plus avoir toujours fréquenté la même école, ne présenter aucun problème majeur d'apprentissage, ni aucun trouble sérieux de la vue ou de l'ouïe.

Les données qui ont été recueillies pour chacun des sujets de l'échantillon sont le sexe, l'âge, le niveau d'études, l'occupation des parents, le rendement général en français et en savoir écouter, parler, lire, écrire et le rendement en mathématiques. Quand il était disponible, le score obtenu à un test du quotient intellectuel a aussi été recueilli.

Il ressort, de l'examen de l'âge des sujets, que cette donnée est constante à l'intérieur d'un même niveau d'études. Ces deux ensembles de données n'en

constituent donc en fait qu'un seul. On a convenu que le *niveau d'études* sera retenu comme variable indépendante.

Les méthodes d'apprentissage de la lecture et de l'orthographe utilisées dans chacune des deux écoles d'où proviennent les sujets de l'échantillon ne sont pas les mêmes. L'une est La méthode dynamique de lecture et de français à l'élémentaire de Guinebretière et al. (1971), l'autre, la Méthode de lecture spontanée de Simone Bussières (1971). Ces deux méthodes adoptent, comme point de départ, une unité plus large que le graphème, soit la phrase ou le mot, ce qui en fait deux méthodes dites globales ou synthétiques. Par ailleurs, à partir de mots ou de phrases, elles incitent à la découverte de la syllabe et de la lettre, ce qui caractérise les méthodes dites analytiques. Ces deux méthodes rallient donc à la fois les approches synthétique et analytique et, de ce fait, peuvent être considérées toutes deux comme des méthodes mixtes. Conséquemment, la méthode d'apprentissage ne sera pas retenue comme variable indépendante mais sera plutôt considérée comme une constante.

Puisque plusieurs études (Sexton, 1965; Rossi, 1968; C.D.S.M., 1970) ont établi que le milieu social influence le rendement scolaire, on a tenté de neutraliser l'influence de ce facteur pour que les étudiants considérés comme forts ou

faibles dans chacune des écoles soient de niveau comparable d'une école à l'autre. Aussi, les deux écoles d'où proviennent les sujets de l'échantillon ont-elles été choisies dans des milieux de niveau socio-économique aussi semblable que possible. Pour vérifier la ressemblance des milieux, on s'est référé, d'abord, de façon informelle, à l'allure générale des quartiers dans lesquels étaient situées les deux écoles. Ensuite, on a examiné la seule donnée disponible dans les dossiers des étudiants, soit l'occupation professionnelle des parents. Or, il ressort, de l'examen de cette information, que les enfants des deux écoles appartiennent effectivement à des milieux semblables. Conséquemment, le niveau socio-économique sera donc considéré comme une constante.

Des données portant sur le rendement général en français et sur celui des quatre savoirs ainsi que sur le rendement en mathématiques ont été recueillies. Si ces données étaient fortement corrélées entre elles, une seule pourrait être retenue pour les fins de cette étude. Or, il ressort d'un test statistique de corrélation de rang[1], le test de Spearman, qui met en relation le rendement général en

---

(1) On a utilisé un test de corrélation de rang plutôt qu'un test de corrélation calculé à partir de données brutes puisque les données portant sur le rendement en français et les divers autres rendements ne proviennent pas de tests standardisés mais de notes attribuées par différents professeurs, à partir de critères qui peuvent varier d'un professeur à l'autre.

français et les autres rendements (Tableau 2.08), que la corrélation entre le rendement général en français et les rendements dans chacun des quatre savoirs et en mathématiques est très forte ($P < 0,0001$). Parce qu'ils sont intercorrélés, un seul de ces rendements peut donc représenter tous les autres. Conséquemment, on peut retenir le *rendement général en français* comme variable indépendante.

Les données recueillies sur le rendement général en français des sujets proviennent de leur dossier scolaire. Elles sont constituées de notes attribuées par les professeurs. Or, on pourrait penser que ces données seraient plus adéquates si elles provenaient des résultats obtenus par les sujets à divers tests reconnus. Afin de vérifier leur fidélité, les résultats obtenus à l'examen de français préparé par le ministère de l'Education du gouvernement du Québec par les sujets de quatrième, cinquième et sixième années (1) ont été recueillis. Outre la cote brute obtenue à cet examen par chacun des sujets, le Ministère fournit également leur position en termes de centile provincial et régional.

---

(1) Seuls les étudiants de sixième année sont soumis à ces examens. Les résultats des étudiants qui étaient en quatrième et en cinquième années lors de l'expérimentation ont été recueillis lorsqu'ils ont passé ces examens un an ou deux après l'expérimentation.

TABLEAU 2.08

COEFFICIENTS DE CORRÉLATION ENTRE LE RENDEMENT GÉNÉRAL EN FRANCAIS ET LES RENDEMENTS DANS LES QUATRE SAVOIRS ET EN MATHÉMATIQUES

RENDEMENTS DIVERS	COEFFICIENTS DE CORRÉLATION
Savoir écouter	0,8171**
Savoir parler	0,7637**
Savoir lire	0,8551**
Savoir écrire	0,7836**
Mathématiques	0,7013**

** $P < 0,001$

Le test de corrélation de rang de Spearman a été utilisé pour mettre en relation les résultats obtenus à l'examen du Ministère et ceux qui étaient attribués par les professeurs et qui constituent les données de la variable *rendement général en français*. Les coefficients de corrélation obtenus sont présentés au tableau 2.09. Comme on peut le constater, il existe, pour les sujets des niveaux 4, 5 et 6, une forte corrélation entre le rang obtenu à l'examen ministériel et celui qui est attribué par les professeurs dans leur école. Ce dernier reflète donc probablement assez fidèlement la compétence des sujets en français. Si c'est vrai pour les étudiants de quatrième, cinquième et sixième années, ce l'est aussi probablement pour ceux de deuxième et de troisième années. On a donc de bonnes raisons de croire que les données qui constituent la variable *rendement général en français* sont pertinentes.

En 1932, Lowry écrivait: "An I.Q. made on a group test may be largely a reading score". L'année suivante, Durrel (1933) arrivait à la même conclusion à la suite d'une étude dans laquelle avaient été impliqués plus de mille enfants. Subséquemment aux travaux de Lowry et Durell, de nombreuses études dans lesquelles le quotient intellectuel

TABLEAU 2.09

COEFFICIENTS DE CORRÉLATION ENTRE LE RENDEMENT GÉNÉRAL EN FRANCAIS ET LES RÉSULTATS A L'EXAMEN MINISTÉRIEL DE FRANCAIS

MESURES DU RENDEMENT A L'EXAMEN GOUVERNEMENTAL	COEFFICIENTS DE CORRÉLATION
Cote brute	0,5894**
Centile provincial	0,6464**
Centile régional	0,5942**

** $P < 0,001$

et le rendement en lecture ont été mis en relation ont été entreprises. Quelques-uns des tests d'intelligence qui ont été ainsi étudiés sont le Pintner General Ability Test (Lennon, 1950), le Henmon-Nelson Intelligence Test (Bevington, 1958), le Stanford-Binet Intelligence Test (Minkler, 1959), le Lorge-Thorndike Intelligence Test (Hage, Stroud, 1959), le California Short Form Test of Mental Ability (Hoyt, Blackmore, 1960), le Peabody Picture Vocabulary Test (Reger, 1962), le Wechsler Intelligence Scale for Children (Bliesmer, 1955). Toutes ces études arrivent à des résultats similaires: le quotient intellectuel et le rendement en lecture sont fortement corrélés.

Puisqu'il a été établi que le rendement en savoir lire est fortement corrélé au rendement général en français (Tableau 2.08), celui-ci devrait être corrélé avec le quotient intellectuel. Les coefficients de corrélation de Spearman obtenus à la suite de la mise en relation de ces deux variables le confirment (Tableau 2.10). On possède ainsi une confirmation supplémentaire sur la pertinence de la variable *rendement général en français*. Quant aux données qui ont été recueillies sur le quotient intellectuel des sujets de l'échantillon, elles ne seront pas utilisées puisqu'elles sont fortement corrélées avec le rendement général en français qui est déjà retenu comme variable indépendante.

TABLEAU 2.10

COEFFICIENTS DE CORRÉLATION ENTRE LE RENDEMENT GÉNÉRAL EN FRANCAIS ET LE QUOTIENT INTELLECTUEL

MESURES DU RENDEMENT EN FRANCAIS	COEFFICIENTS DE CORRÉLATION
Cote brute	0,6621**
Centile provincial	0,4362**
Centile régional	0,4438**

** P < 0,001

Pour déterminer les catégories de rendement en français, la moitié des étudiants de chaque niveau et de chaque école qui avaient les meilleures notes en français ont été classés comme forts. Les autres ont été considérés comme faibles.

Les variables indépendantes retenues pour les fins du présent travail sont donc le *niveau d'études* et le *rendement en français*. Quant aux variables dépendantes qui ont été utilisées pour analyser la production écrite des sujets de l'échantillon, elles seront définies au moment de la présentation et de l'analyse des résultats.

2.4 Déroulement de l'expérience

L'expérience a eu lieu au cours du troisième mois de l'année scolaire. A cause du jeune âge de certains des sujets de l'échantillon, l'épreuve orthographique a été présentée en quatre parties aux étudiants de deuxième et de troisième années. Pour chacune des classes de ces niveaux d'études, les sessions ont eu lieu le matin de quatre jours consécutifs. Les étudiants de quatrième, cinquième et sixième y ont été soumis l'après-midi et en deux fois.

L'expérience s'est déroulée dans le local régulier de classe des élèves et ceux-ci étaient installés à la place qu'ils occupent habituellement. Tous les étudiants de la classe ont participé à l'expérimentation même si certains n'allaient pas être retenus à cause de l'un ou l'autre des critères déjà énoncés au point 2.3[(1)].

On a distribué à chaque étudiant une feuille sur laquelle il écrivait en haut, à gauche, son nom et à droite, un chiffre de un à quatre. Ce chiffre correspondait à la série de logatomes qui allait lui être dictée. Sur la feuille, vingt-six lignes numérotées avaient été tracées.

---

(1) Vingt étudiants de chaque groupe ont été retenus pour fins d'analyse. Quand, après l'application des critères, plus d'étudiants que ce nombre pouvaient être retenus, on a éliminé de façon aléatoire les étudiants en surplus.

Les logatomes devaient y être inscrits. Dans les classes de deuxième et de troisième années, chacune des quatre séances a duré entre vingt et trente minutes. Dans les classes de quatrième à sixième année, chacune des deux séances a duré environ trente minutes.

Le professeur n'était pas présent dans la classe au moment de l'expérimentation. Le test a été présenté par deux personnes, une dans chaque école. La lecture des logatomes avait au préalable été répétée plusieurs fois par les deux expérimentateurs afin que leur prononciation se ressemble le plus possible. Aux élèves plus jeunes (deuxième et troisième années), les logatomes étaient présentés comme des mots très savants qu'ils ne connaissaient probablement pas encore mais qu'ils étaient certainement capables d'écrire correctement s'ils étaient très attentifs. Un tel scénario ne pouvait être utilisé auprès des étudiants des niveaux supérieurs. Aussi leur disait-on qu'on leur présenterait une liste de mots fictifs qu'ils devaient écrire tout comme si ces mots étaient des mots français.

Lors de la lecture des logatomes, chaque phonème était clairement prononcé et chacun des logatomes était

automatiquement lu deux fois. Les logatomes encore mal compris à la deuxième lecture étaient répétés autant de fois que nécessaire. Invariablement les étudiants ont semblé participer de façon très consciencieuse à l'expérimentation.

CHAPITRE III

ANALYSE DES RÉSULTATS

3.1 Introduction

Les recherches pédagogiques sur l'orthographe lexicale qui définissent des vocabulaires de base et les endroits, à l'intérieur des mots, où les erreurs sont le plus souvent commises, ne renseignent pas, finalement, sur la façon dont se fait l'apprentissage de l'orthographe lexicale.

Parce que les études qui ont porté sur l'apprentissage de la lecture ont poussé beaucoup plus loin leurs analyses des mécanismes d'apprentissage de cette habileté et que la lecture et l'écriture présentent plusieurs points de ressemblance, on s'est inspiré des conclusions des études sur la lecture pour énoncer des hypothèses sur la façon don. se fait l'apprentissage de l'orthographe lexicale.

Gibson et al. (1962-63) ont établi que l'unité de perception chez le lecteur expérimenté, est constituée d'une suite de graphèmes qui entretient une relation invariante avec une suite donnée de phonèmes dans une langue déterminée. Des études portant sur la lecture ont également établi que les

lecteurs expérimentés dans une langue connaîtraient, au moins implicitement, les règles de formation des suites graphémiques de cette langue (Gibson et al., 1970).

A la suite des études sur l'apprentissage de la lecture, il ne paraît pas invraisemblable de penser que celui qui apprend à orthographier dans une langue suivrait un processus semblable à celui qui apprend à lire. On a ainsi été amené à énoncer des hypothèses de travail portant sur la façon dont se fait l'apprentissage de l'orthographe lexicale.

Pour vérifier leur justesse, une épreuve orthographique constituée de logatomes a été construite et soumise à des élèves de la deuxième à la sixième année du niveau primaire. La transcription de ces logatomes par chacun des sujets de l'échantillon a été codifiée et placée sur support informatique. L'analyse des suites graphémiques auxquelles les étudiants ont eu recours pour représenter les mots fictifs qui leur ont été dictés permettra de vérifier les quatre hypohèses de travail qui ont été suggérées. Ces analyses seront faites en tenant compte des variables indépendantes retenues pour les fins de cette étude, soit le *niveau d'études* auquel un étudiant est inscrit et son *rendement en français*.

3.2 Vérification de l'hypothèse E1

HYPOTHÈSE E1 : L'apprentissage de l'orthographe lexicale ne se fait pas seulement par la mémorisation systématique de la représentation graphémique de chacun des mots de la langue.

Il existe, en français, des règles de correspondance entre suites phonologiques et suites graphémiques. L'ensemble de ces règles constitue le code phono-graphique du français. Celui qui a assimilé ce code et à qui on demande d'écrire des logatomes qui lui sont présentés comme des mots français peut alors respecter ces règles de correspondance. On peut donc conjecturer qu'une personne qui apprend à orthographier le français fasse montre d'une maîtrise progressive de ces règles de sorte que le code qu'elle utilise se conformerait de plus en plus au code phono-graphique du français même dans une tâche de reproduction graphémique de logatomes. Si tel est le cas, l'apprentissage de l'orthographe lexicale ne se ferait donc pas par la seule mémorisation de chacun des mots de la langue. Pour vérifier cette hypothèse, on a défini une variable dépendante: l'indice de similitude au code français.

3.2.1 L'indice de similitude au code français

Basée sur les fréquences des expressions graphémiques théoriquement utilisées en français pour représenter les phonèmes /s/, /c/ et /o/, l'indice de similitude au code français sert à mesurer le degré de rapprochement du code d'un individu au code français. Pour apprécier statistiquement la différence entre le code d'un individu et le code français, on a utilisé le test du khi deux dont la formule est :

$$\chi^2 = \sum \frac{(fo - f_t)^2}{f_t}$$

χ^2 = khi deux

fo = fréquences observées ou empiriques

f_t = fréquences attendues ou théoriques

Cette statistique fournit en effet une mesure de l'écart qui existe entre des fréquences empiriques et des fréquences théoriques. L'écart entre les deux codes mesuré par le khi deux est toutefois difficile à évaluer puisque cette statistique n'a pas de valeur maximum. Elle a donc été transformée en coefficient de corrélation dont les valeurs sont comprises entre zéro et un. La formule suivante utilisée pour effectuer cette transformation est celle utilisée par Spiegel (1961, p. 204):

$$r = \sqrt{\frac{\chi^2}{N\,(k-1)}}$$

r = coefficient de corrélation

χ^2 = khi deux

N = nombre d'occurrences dans l'épreuve orthographique des phonèmes étudiés

k = nombre d'expressions graphémiques retenues pour le calcul du χ^2

Si les fréquences des expressions graphémiques utilisées par un individu pour représenter les phonèmes étudiés se rapprochent des fréquences du code français, la valeur du khi deux sera faible. Conséquemment, le coefficient de corrélation tendra vers zéro. Il tendra par contre vers un si l'écart entre le code d'un individu et le code français est grand.

3.2.1.1 L'indice de similitude au code français pour le phonème /s/.

Les indices de similitude au code français pour le phonème /s/ présentés en fonction des variables indépendantes *niveau d'études* et *rendement en français* apparaissent au tableau 3.01. Si on considère d'abord les résultats présentés en fonction de la variable *niveau d'études*, il ressort de l'examen de ces données que la corrélation entre le code de: sujets de l'échantillon et le code français est de plus en plus forte de la deuxième à la sixième année. Le progrès le plus prononcé se produit entre la deuxième et la troisième année alors que l'indice passe de 0,45 à 0,36. Le progrès le plus faible se situe entre la cinquième et la sixième année alors que l'indice passe de 0,28 à 0,26.

Si on considère maintenant les résultats présentés en fonction de la variable *rendement en français*, on remarque que dès la deuxième année, les étudiants considérés comme forts en français présentent une avance marquée sur les faibles dans l'acquisition du code français puisque leur indice de similitude est de 0,42 par rapport à celui des faibles qui est de 0,48. En général, les forts ont environ une année d'avance sur les faibles. C'est ainsi par exemple que les forts de troisième année ont un code qui se rapproche davantage du code français que les faibles de quatrième année, l'indice de similitude au code français étant, pour ces deux groupes, de 0,30 et 0,35.

TABLEAU 3.01

INDICES DE SIMILITUDE AU CODE FRANCAIS
POUR LE PHONÈME /s/ EN FONCTION DU
NIVEAU D'ÉTUDES ET DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS	FAIBLES	MOYENNE
2	0,42	0,48	0,45
3	0,31	0,41	0,36
4	0,30	0,35	0,33
5	0,27	0,29	0,28
6	0,24	0,28	0,26
Moyenne	0,31	0,36	0,34

Le tableau 3.02 qui présente les résultats d'une analyse de variance effectuée à partir de l'indice de similitude au code français pour le phonème /s/ confirme l'effet des variables indépendantes *niveau d'études* et *rendement en français* sur le degré de rapprochement entre le code des individus et le code français puisque l'effet de ces variables est significatif au niveau de 0,001. Cependant, bien que l'écart entre les indices de similitude au code français des forts et des faibles tende à décroître à mesure que le niveau s'élève, l'interaction *niveau d'études - rendement en français* n'est pas significative, le niveau de signification, égal à 0,557, étant plus grand que le seuil de confiance établi à 0,05. Autrement dit, l'absence d'interaction significative indique que l'effet de la variable *niveau d'études* ne varie pas de façon significative d'une catégorie de rendement en français à l'autre. De la même façon, l'effet de la variable *rendement en français* n'est pas significativement différent d'un niveau d'études à l'autre.

TABLEAU 3.02

RÉSULTATS DE L'ANALYSE DE VARIANCE EFFECTUÉE À PARTIR DE L'INDICE DE SIMILITUDE AU CODE FRANCAIS POUR LE PHONÈME /s/

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	2078,0	4	17,3	0,001
Rendement en français	1262,9	1	10,5	0,001
Interaction niv.-rend.	90,4	4	0,8	0,557
Résidus	119,9	183		

3.2.1.2 L'indice de similitude au code français pour le phonème /ɛ/

Les indices de similitude au code français pour le phonème /ɛ/ présentés en fonction des variables indépendantes *niveau d'études* et *rendement en français* apparaissent au tableau 3.03. On remarque, à l'examen des résultats présentés en fonction de la variable *niveau d'études*, que de la deuxième à la cinquième année, le code des individus de l'échantillon se rapproche progressivement du code français. Tandis que pour le phonème /s/, on observait une relative stabilisation du code entre la cinquième et la sixième année, pour le phonème /ɛ/, il semble se produire un léger recul en sixième année. Il ressort également de ces données que les progrès les plus marqués dans le rapprochement entre le code des individus et le code français se produisent, comme pour le phonème /s/, entre la deuxième et la troisième année alors que l'indice passe de 0,42 à 0,29.

Si on considère les résultats présentés en fonction de la variable *rendement en français*, on remarque que de la deuxième à la quatrième année, l'écart entre les indices de similitude des étudiants forts et faibles en français est marqué. L'avance des étudiants forts sur les étudiants faibles est toutefois de moins d'un an sauf pour les forts de troisième année dont le code se rapproche davantage du code français que celui des étudiants faibles de quatrième année. Les

TABLEAU 3.03

INDICES DE SIMILITUDE AU CODE FRANCAIS
POUR LE PHONÈME /ɛ/ EN FONCTION DU
NIVEAU D'ÉTUDES ET DU RENDEMENT
EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS	FAIBLES	MOYENNE
2	0,39	0,45	0,42
3	0,27	0,31	0,29
4	0,26	0,35	0,31
5	0,20	0,21	0,21
6	0,21	0,22	0,22
Moyenne	0,26	0,31	0,29

indices de similitude au code français pour ces deux groupes sont en effet de 0,27 et de 0,35. Il est étonnant de constater qu'entre la troisième et la quatrième année, il n'y a pas de progrès ni chez les forts, ni chez les faibles alors que l'indice se stabilise chez les forts et qu'il régresse de façon assez importante chez les faibles. A partir de la cinquième année, l'écart entre les forts et les faibles devient pratiquement nul.

Une analyse de variance effectuée à partir de l'indice de similitude au code français pour le phonème /ɛ/ et dont les résultats sont présentés au tableau 3.04 confirme l'effet des variables indépendantes sur le taux de variation de l'indice de similitude. Le niveau de signification correspondant à la variable *niveau d'études* est en effet de 0,001; celui qui correspond à la variable *rendement en français* est de 0,021. Les étudiants forts ont donc un code dont le degré de similitude au code français est significativement plus élevé que celui des faibles. De même, le niveau d'études joue clairement un rôle dans le degré de rapprochement du code français.

TABLEAU 3.04

RÉSULTATS DE L'ANALYSE DE VARIANCE EFFECTUÉE À PARTIR DE L'INDICE DE SIMILITUDE AU CODE FRANCAIS POUR LE PHONÈME /ɛ/

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	2659,5	4	15,3	0,001
Rendement en français	940,4	1	15,4	0,021
Interaction niv.-rend.	115,4	4	0,7	0,618
Résidus	173,8	183		

3.2.1.3 L'indice de similitude au code français pour le phonème /o/

Les indices de similitude au code français pour le phonème /o/ présentés en fonction des variables indépendantes *niveau d'études* et *rendement en français* apparaissent au tableau 3.05. A l'examen des données présentées en fonction de la variable *niveau d'études*, on constate que de la deuxième à la cinquième année, il y a un rapprochement graduel entre le code des individus et le code français avec, comme pour les phonèmes /s/ et /ɛ/, un saut très prononcé entre la deuxième et la troisième année. Les indices de similitude au code passent en effet, pour ces niveaux d'études, de 0,57 à 0,39. Entre la cinquième et la sixième année, on remarque par contre une régression, cette fois beaucoup plus importante que celle qui a été observée pour le phonème /ɛ/ puisque la valeur des indices passe de 0,34 à 0,40 alors qu'elle ne passait que de 0,21 à 0,22 pour le phonème /ɛ/.

Si on considère maintenant les résultats présentés en fonction de la variable *rendement en français*, on constate que tout comme pour les phonèmes /s/ et /ɛ/, les étudiants qui ont un rendement supérieur en français ont un code qui se rapproche clairement plus du code français que les étudiants dont le rendement est plus faible. Les forts de troisième année dont l'indice de similitude au code français a une valeur de 0,34 ont même une avance de plus d'un an sur

TABLEAU 3.05

INDICES DE SIMILITUDE AU CODE FRANCAIS
POUR LE PHONÈME /ɔ/ EN FONCTION DU
NIVEAU D'ÉTUDES ET DU RENDEMENT
EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS	FAIBLES	MOYENNE
2	0,54	0,60	0,57
3	0,34	0,43	0,39
4	0,35	0,37	0,36
5	0,31	0,36	0,34
6	0,35	0,45	0,40
Moyenne	0,38	0,44	0,41

les étudiants faibles puisque les faibles de quatrième année ont un indice dont la valeur est de 0,37. Quant aux étudiants forts de quatrième année, ils ont une avance d'un an sur les faibles puisque leur indice est équivalent à celui du groupe faible de cinquième année, leurs valeurs étant respectivement de 0,35 et 0,36. Entre la cinquième et la sixième année, on observe, chez les forts et les faibles, une régression assez importante. Chez les forts, l'indice de similitude au code passe en effet de 0,31 à 0,35, ce qui représente un retour au niveau de la quatrième année puisque l'indice qui correspond à ce niveau est de 0,35. Chez les faibles, la régression est encore plus importante. L'indice passe de 0,36 à 0,45, ce qui représente un retour au niveau de la troisième année à laquelle correspond en effet un indice de 0,43.

Il ressort des résultats de l'analyse de variance effectuée à partir de l'indice de similitude au code français pour le phonème /o/ que l'effet des variables indépendantes *niveau d'études* et *rendement en français* sur le degré de rapprochement entre le code des individus et le code français est significatif comme l'indique, au tableau 3.06, le niveau de signification correspondant à ces variables.

TABLEAU 3.06

RÉSULTATS DE L'ANALYSE DE VARIANCE EFFECTUÉE À PARTIR DE L'INDICE DE SIMILITUDE AU CODE FRANÇAIS POUR LE PHONÈME /o/

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	3171,8	4	10,4	0,001
Rendement en français	2145,3	1	7,0	0,009
Interaction niv.-rend.	80,6	4	0,3	0,901
Résidus	304,9	183		

3.2.2 Conclusion

Selon l'hypothèse E1, l'apprentissage de l'orthographe lexicale ne se ferait pas seulement par la mémorisation systématique de la représentation graphémique de chacun des mots de la langue. Pour vérifier cette hypothèse, on a dicté des logatomes à des étudiants du niveau primaire. Le recours à des logatomes présente l'avantage de neutraliser l'effet de la mémorisation des mots sur la production écrite des étudiants. Le dépouillement de cette production a permis de mesurer, pour chacun des phonèmes étudiés, le degré de rapprochement des fréquences des expressions graphémiques utilisées par un individu des fréquences de ces expressions dans la langue française. Le test de logatomes a en effet été construit de façon à induire la production de graphies dans les mêmes proportions que celles dans lesquelles elles apparaissent dans la langue française.

On a constaté, à l'examen des résultats, que les proportions dans lesquelles les étudiants utilisent les diverses expressions graphémiques pour représenter un phonème se rapprochent progressivement de celles du français à mesure que les étudiants accèdent à des niveaux de scolarité plus avancés. De plus, le degré de rapprochement entre le code des individus et le code français est significativement plus élevé chez les étudiants dont le rendement en français est supérieur que chez ceux dont le rendement est plus faible. Ces

résultats indiquent qu'il y aurait un apprentissage quelconque de l'orthographe lexicale qui ne serait clairement pas attribuable à la mémorisation. L'hypothèse E1 est donc confirmée.

Il est intéressant de remarquer que, des trois phonèmes étudiés, c'est-à-dire /s/, /ɛ/ et /o/, c'est pour le phonème /ɛ/ que le degré de rapprochement entre le code de l'ensemble des sujets de l'échantillon et le code français est le plus élevé, sa valeur moyenne étant de 0,29 (tableau 3.03). Le phonème /s/ arrive en seconde position avec un degré de rapprochement moyen égal à 0,34 (tableau 3.01). Enfin, le phonème /o/, avec un degré de rapprochement moyen de 0,41 (tableau 3.05), se classe au dernier rang.

Pour chacun des phonèmes, les étudiants forts de chaque niveau d'études ont, par rapport aux étudiants faibles, un code qui se rapproche davantage du code français. Pour le phonème /s/, cette avance est d'environ un an pour chaque niveau d'études. Pour le phonème /o/, on remarque également une avance d'environ un an des forts sur les faibles mais seulement à partir de la troisième année. Pour le phonème /ɛ/, les forts de troisième année ont une avance de plus d'un an sur les faibles. Ceux-ci rejoignent toutefois les forts en cinquième année.

Si on ne considère que la variable *niveau d'études*, on remarque qu'il se produit, pour chacun des phonèmes étudiés, un saut important entre la deuxième et la troisième année dans le degré de rapprochement entre le code des sujets de l'échantillon et le code français. Ce progrès pourra peut-être être expliqué ultérieurement lors de la vérification des trois autres hypothèses de travail. Une autre remarque intéressante qui découle de l'examen des résultats compilés pour la vérification de l'hypothèse E1 est la stabilisation dans le cas des phonèmes /s/ et /ɛ/ et la régression dans le cas du phonème /o/ du degré de rapprochement du code des individus de sixième année par rapport à celui des individus de cinquième année. Ce phénomène pourrait peut-être s'expliquer par le fait que plus le niveau d'études est élevé, plus l'apprentissage de l'orthographe lexicale perdrait de l'importance dans l'apprentissage du langage écrit. Or si, au moment où il passe au second plan, cet apprentissage n'a pas déjà été suffisamment intégré, il n'est pas étonnant de remarquer une régression au niveau de la compétence des étudiants en orthographe lexicale. C'est donc peut-être trop tôt que l'on diminue son importance dans l'apprentissage du langage écrit.

3.3 Vérification de l'hypothèse E2

HYPOTHÈSE E2 : L'individu qui fait l'apprentissage de l'orthographe lexicale française apprend progressivement à associer, à chaque phonème, les diverses expressions graphémiques qui peuvent le représenter.

On a démontré que l'apprentissage de l'orthographe lexicale ne se fait pas par la seule mémorisation de la représentation graphémique de chacun des mots de la langue. Par ailleurs, on sait qu'en français, à de nombreux phonèmes correspondent plusieurs expressions graphémiques. L'association de ces diverses graphies au phonème approprié constituerait une composante de l'acquisition du code phonographique français et par conséquent, de l'orthographe lexicale.

Pour vérifier cette hypothèse, on a d'abord développé un premier outil, l'indice d'étendue du code, qui mesure le nombre d'expressions graphémiques correctes différentes auxquelles les sujets de l'échantillon ont eu recours pour représenter, dans l'épreuve orthographique, les phonèmes /s/, /ɛ/ et /o/. Pour obtenir des indications supplémentaires sur la connaissance que les sujets de l'échantillon ont des expressions graphémiques qui peuvent être associées à un phonème, deux outils de mesure complémentaires ont été construits. L'un d'eux, l'indice de popularité des graphies,

présente le pourcentage d'élèves de chaque groupe qui a utilisé les différentes graphies correctes pour représenter chacun des phonèmes étudiés. L'autre outil, l'indice d'utilisation des graphies, reflète la proportion d'utilisation, par les sujets de l'échantillon, des principales expressions graphémiques qui peuvent représenter l'un des phonèmes étudiés. Les informations fournies par ces deux outils permettront de savoir, entre autres, si les expressions graphémiques que les individus ont appris à associer à chaque phonème sont les mêmes chez les étudiants forts et les étudiants faibles de même que d'un niveau d'études à l'autre.

3.3.1 L'indice d'étendue du code

Le tableau 3.07 présente, pour les phonèmes /s/, /ɛ/ et /o/, l'indice d'étendue du code, c'est-à-dire le nombre moyen d'expressions graphémiques pertinentes utilisées par les individus de chacun des niveaux d'études et de chacune des catégories de rendement en français.

En moyenne, les sujets de l'échantillon ont utilisé 4,3 expressions graphémiques différentes pour représenter le phonème /s/. Ce nombre varie très peu d'un niveau d'études à l'autre de même que pour chacune des catégories de rendement en français.

L'examen des expressions graphémiques utilisées par chacun des sujets de l'échantillon pour représenter le phonème /s/ révèle toutefois quelques cas particuliers. En effet, sur l'ensemble des individus qui ont été soumis à l'épreuve orthographique, six étudiants ont obtenu un indice d'étendue très faible. Alors qu'un étudiant n'a utilisé que deux expressions graphémiques différentes, cinq n'en ont utilisé que trois pour représenter les soixante-dix-sept occurrences du phonème /s/. Il ne semble pas y avoir de relation entre le niveau d'études ou la catégorie de rendement en français et la faible étendue du code, ces six étudiants provenant de niveaux d'études différents ainsi que des deux catégories

TABLEAU 3.07

INDICE D'ÉTENDUE DU CODE
PRÉSENTÉ EN FONCTION DU NIVEAU
D'ÉTUDES ET DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS	PHONÈMES		
		/s/	/ɛ/	/o/
2	Forts	4,2	4,5	2,5
	Faibles	4,0	4,0	1,9
3	Forts	4,1	5,1	3,4
	Faibles	4,4	4,7	2,9
4	Forts	4,5	5,8	3,1
	Faibles	4,4	5,2	3,3
5	Forts	4,3	5,3	3,6
	Faibles	4,7	4,9	3,3
6	Forts	4,4	5,1	3,3
	Faibles	4,2	5,2	3,7
Moyenne		4,3	5,0	3,1

de rendement en français.

Les expressions graphémiques utilisées pour représenter le phonème /ɛ/ sont, en moyenne, au nombre de 5,0. Ceux qui en ont utilisé le moins sont les écoliers faibles de deuxième année. Il faut noter qu'ils en ont tout de même utilisées 4,0. Ceux qui en ont utilisé le plus sont les forts de quatrième année. Une moyenne de 5,8 expressions graphémiques différentes ont été utilisées par ce groupe. En général, le nombre moyen de graphies pour les divers groupes s'écarte assez peu de la moyenne globale.

On remarque toutefois, à l'examen des expressions graphémiques utilisées par chacun des sujets de l'échantillon pour représenter le phonème /ɛ/ que trois étudiants ont un code très restreint. Ces étudiants n'ont en effet utilisé que deux ou trois expressions graphémiques différentes pour représenter les soixante-dix-sept occurrences du phonème /ɛ/. Comme pour le cas du phonème /s/, les étudiants qui ont un code plus restreint ne sont pas concentrés dans un seul niveau d'études ou une seule catégorie de rendement en français.

Pour représenter le phonème /o/, les sujets de l'échantillon n'ont utilisé, en moyenne, que 3,1 graphies différentes. Ceux qui en ont utilisé le moins sont les

étudiants forts et faibles de deuxième année qui n'ont eu recours qu'à 2,5 et 1,9 expressions graphémiques différentes pour représenter les cinquante-six occurrences de ce phonème.

De l'examen des expressions graphémiques utilisées par chacun des sujets de l'échantillon pour représenter le phonème /o/, il ressort que contrairement aux cas des phonèmes /s/ et /ɛ/, les étudiants qui ont utilisé un nombre particulièrement restreint de graphies différentes sont concentrés surtout en deuxième année. Puisqu'ils sont assez nombreux, soit neuf, on comprend mieux pourquoi l'indice d'étendue du code à ce niveau est plus faible. Il n'y a cependant pas qu'à ce niveau qu'on retrouve des codes restreints puisque neuf autres étudiants répartis entre les autres niveaux d'études n'ont également utilisé qu'une ou deux expressions graphémiques différentes pour représenter le phonème /o/.

Des données présentées au tableau 3.07, il semble qu'on puisse déduire que dès la deuxième année, les sujets de l'échantillon associent déjà un assez bon nombre d'expressions graphémiques pertinentes aux phonèmes étudiés. Pour les phonèmes /s/ et /ɛ/, les étudiants de deuxième année réussissent en effet cette tâche aussi bien que ceux des niveaux supérieurs. Pour le phonème /o/, cette

performance est atteinte dès la troisième année. L'apprentissage de l'association des diverses expressions graphémiques aux phonèmes qu'elles représentent se ferait donc très tôt, vraisemblablement au cours des deux premières années d'études.

On aurait pu croire que la distribution des fréquences, en français, des diverses expressions graphémiques qui représentent les phonèmes /s/ (tableau 2.01), /ɛ/ (tableau 2.02) et /o/ (tableau 2.03) puisse avoir une influence sur le nombre d'expressions graphémiques différentes utilisées par les divers groupes d'élèves. En effet, la valeur relativement faible de l'indice d'étendue du code pour représenter le phonème /o/ (3,1), aurait pu s'expliquer par le fait que ce phonème est représenté, en français, dans plus de 80% des cas, par seulement deux expressions graphémiques différentes, soit *au* et *o*. A cause de leur fréquence élevée, ce sont ces graphies que les sujets de l'échantillon auraient apprises d'abord. Progressivement, ils se seraient rendu compte que d'autres expressions graphémiques peuvent parfois représenter le phonème /o/ et auraient eu recours à une ou deux d'entre elles.

Cependant, quoique le phonème /ɛ/ soit représenté, en français, dans plus de 70% des cas, par les deux seules

expressions graphémiques *e* et *aɩ*, l'indice d'étendue du code pour ce phonème, qui a une valeur de 5,0, est sensiblement plus élevé que celui du phonème /o/ dont la valeur est de 3,1. Il ne semble donc pas que ce soit la distribution des fréquences théoriques en français des diverses expressions graphémiques qui représentent les phonèmes étudiés qui puisse expliquer la différence dans la valeur des indices d'étendue du code.

L'examen des graphies utilisées par les sujets de l'échantillon dans l'épreuve orthographique permettra peut-être de comprendre cette différence. Cet examen des graphies se fera au moyen des indices de popularité et d'utilisation des expressions graphémiques pour représenter les phonèmes /ɛ/, /ɛ/ et /o/.

3.3.2 Les indices de popularité et d'utilisation des graphies

L'indice d'étendue du code indique combien de graphies pertinentes sont utilisées en moyenne par les individus d'un groupe. Pour mieux apprécier les comportements qui expliquent cette moyenne, on examine, dans ce qui suit, la popularité des expressions graphémiques, c'est-à-dire le pourcentage des individus d'un groupe qui les utilisent ainsi que les proportions dans lesquelles chacune est utilisée par les sujets de l'échantillon.

3.3.2.1 L'indice de popularité des graphies du phonème /s/

Les proportions d'étudiants forts et faibles de chaque niveau d'études et de chaque catégorie de rendement en français qui ont utilisé les diverses expressions graphémiques du phonème /s/ sont présentées au tableau 3.08. Selon l'indice d'étendue du code, un même nombre d'expressions graphémiques différentes a été utilisé, en moyenne, par les étudiants forts et faibles de chacun des niveaux d'études du cours primaire. Il ressort toutefois, de l'indice de popularité des graphies, que plus d'étudiants de troisième année que de deuxième année ont utili-

TABLEAU 3.08

POURCENTAGE D'ÉTUDIANTS DE CHAQUE NIVEAU D'ÉTUDES ET DE CHAQUE CATÉGORIE DE RENDEMENT EN FRANCAIS QUI ONT UTILISÉ LES DIVERSES EXPRESSIONS GRAPHÉMIQUES DU PHONÈME /s/

NIVEAUX	RENDEMENTS	EXPRESSIONS GRAPHÉMIQUES						
		s %	ss %	c %	ç %	t %	sc %	autres %
2	Forts	100	89	100	6	89	28	22
	Faibles	100	76	100	12	71	23	0
3	Forts	100	85	95	60	50	25	0
	Faibles	100	85	85	60	65	35	10
4	Forts	100	100	100	55	80	10	0
	Faibles	100	85	100	50	70	25	5
5	Forts	100	100	100	80	35	10	0
	Faibles	100	100	100	55	75	45	5
6	Forts	100	100	100	70	35	10	0
	Faibles	100	95	100	55	75	25	0
FRÉQUENCES THÉORIQUES		32,4	32,0	22,7	6,5		6,4	

sé plus de graphies différentes pour représenter le phonème /s/. Par contre, de la troisième à la sixième année, un nombre à peu près constant d'étudiants a utilisé les diverses graphies de ce phonème. On remarque de plus que la presque totalité des sujets de l'échantillon a utilisé au moins une fois, dans l'épreuve orthographique, les graphies du phonème /s/ auxquelles correspondent une fréquence théorique élevée soit *s*, *ss* et *c*. Peu d'étudiants de deuxième année se sont servis de la graphie *ç*. A l'exception des groupes de forts de cinquième et de sixième années dont respectivement 80% et 70% des individus se sont servis de cette graphie, environ la moitié seulement des étudiants des autres groupes l'ont utilisée.

Il est intéressant de remarquer que la graphie *ç* à laquelle correspond une fréquence théorique de 6.5% (tableau 2.01) semble moins connue par les sujets de deuxième année que la grahie *t* qui est pourtant plus rare. Alors que seulement 6% des étudiants forts et 12% des étudiants faibles de ce niveau d'études utilisent la graphie *ç*, 89% des forts et 71% des faibles ont recours à la graphie *t*.

Il est également intéressant de remarquer que pour les expressions graphémiques qui représentent rarement le phonème /s/, soit *t* et *sc*, on observe un type de distribution semblable des étudiants. En effet, à l'exception

des étudiants de deuxième année chez qui il n'y a pas vraiment de différence entre les forts et les faibles, à presque tous les autres niveaux, sensiblement moins d'étudiants dont le rendement en français est supérieur utilisent ces graphies. De plus, à peu près le même pourcentage d'étudiants forts et faibles de deuxième année et d'étudiants faibles de chacun des autres niveaux d'études les utilisent. L'examen des données suggère que dès la troisième année, plusieurs étudiants forts se seraient rendu compte que certaines expressions graphémiques ne représentent que rarement le phonème /s/. Confrontés au problème de la réalisation graphique de ce phonème dans un logatome, ils auraient plutôt eu recours aux expressions graphémiques les plus fréquentes. Quant aux étudiants faibles de chacun des niveaux d'études, ils n'auraient peut-être pas pris conscience du fait que certaines graphies ont une fréquence plus faible que d'autres. A moins que la reproduction d'un mot aussi rare qu'un logatome ne les ait incités à recourir justement aux graphies les plus rares.

L'indice de popularité des graphies du phonème /s/ renseigne sur le pourcentage d'étudiants de chaque niveau d'études et de chaque catégorie de rendement en français qui ont utilisé au moins une fois, dans l'épreuve orthographique, les expressions graphémiques qui représentent ce

phonème. L'examen des données fournies par l'indice d'utilisation des graphies du phonème /s/ permettra de savoir de façon plus précise combien chacune des expressions graphémiques a été utilisée pour réaliser les soixante-dix-sept occurrences de ce phonème.

3.3.2.2 L'indice d'utilisation des graphies du phonème /s/.

Les proportions dans lesquelles les principales expressions graphémiques qui représentent le phonème /s/ en position médiane ont été utilisées par les sujets de l'échantillon sont présentées au tableau 3.09. Il ressort de ces données que chaque groupe d'étudiants sur-utilise la graphie ś. En effet, alors que selon le code standard, cette graphie devrait être utilisée dans une proportion de 32,4%, les différents groupes d'étudiants l'utilisent dans des proportions qui varient de 39,4% à 65,9%. Ceux qui se rapprochent le plus de la fréquence théorique sont les étudiants forts de troisième année; ceux qui s'en éloignent le plus sont les étudiants forts de deuxième année suivis de très près par les étudiants faibles du même niveau. En fait, avec des proportions d'utilisation de 65,9% et de 64,3%, les étudiants forts et faibles de deuxième année utilisent la graphie ś deux fois plus souvent qu'ils ne devraient. Un tel écart entre la fréquence théorique de la graphie ś et sa fréquence d'utilisation par les sujets de deuxième année ainsi que le rapprochement des fréquences théorique et empirique chez les étudiants forts de troisième année contribuent à expliquer le saut qui a été observé, lors de la vérification de l'hypothèse E1, dans la valeur de l'indice de similitude pour le phonème /s/ entre la deuxième et la troisième année.

On aurait pu croire que la sur-utilisation de la gra-

TABLEAU 3.09

POURCENTAGES D'UTILISATION DES EXPRESSIONS GRAPHÉMIQUES LES PLUS FRÉQUENTES POUR REPRÉSENTER LE PHONÈME /s/

NIVEAUX	RENDEMENTS	EXPRESSIONS GRAPHÉMIQUES					
		s %	ss %	c %	ç %	autres bonnes %	autres mauvaises %
2	Forts	65,9	6,7	23,8	0,1	2,9	0,6
	Faibles	64,3	5,0	25,6	0,2	3,7	1,2
3	Forts	39,4	20,7	31,0	6,4	1,7	0,8
	Faibles	49,9	8,5	33,5	4,5	2,5	1,1
4	Forts	41,5	21,7	30,3	4,0	1,9	0,6
	Faibles	52,5	13,0	28,2	2,0	2,7	1,6
5	Forts	41,9	21,9	29,4	4,3	1,6	0,9
	Faibles	51,4	19,0	22,2	2,7	3,3	1,2
6	Forts	43,8	27,2	23,3	3,8	1,3	0,6
	Faibles	47,2	27,0	20,0	3,1	2,1	0,6
Fréq. moyennes	Forts	46,5	19,6	27,6	3,7	1,9	0,7
	Faibles	53,1	14,5	25,9	2,5	2,9	1,1
Fréquences théoriques		32,4	32,0	22,7	6,5	6,4	0,0

phie ȣ pouvait être liée au fait que cette graphie est fréquemment utilisée en français standard écrit. Or, il ressort des données présentées au tableau 3.09, que tous les groupes d'étudiants sous-utilisent l'expression graphémique ȣȣ dont la fréquence théorique, qui est de 32,0%, est à peu près la même que celle qui correspond à la graphie ȣ, c'est-à-dire 32,4%. L'expression graphémique ȣȣ est en effet utilisée dans des proportions aussi faibles que 6,7%, 5,0% et 8,5% par les étudiants forts et faibles de deuxième année et les étudiants faibles de troisième année. Même pour ceux qui l'utilisent le plus souvent, soit les étudiants forts et faibles de sixième année, les proportions d'utilisation qui sont de 27,2% et 27,0% sont inférieures à la fréquence théorique. On note toutefois qu'à ce niveau, les fréquences empiriques se rapprochent sensiblement de la fréquence théorique.

Parce qu'à tous les niveaux d'études, il y a sur-utilisation de la graphie ȣ et sous-utilisation de l'expression graphémique ȣȣ à laquelle correspond une fréquence théorique équivalent à celle de ȣ, on peut mettre en cause l'influence de la fréquence théorique d'une graphie sur sa fréquence d'utilisation.

Outre l'expression graphémique ȣȣ, une seconde graphie est sous-utilisée par chacun des groupes d'étudiants. Il

s'agit de la graphie ç qui n'est pratiquement pas utilisée par les étudiants de deuxième année et qui l'est dans une proportion de 6,4% par ceux qui l'ont utilisée le plus souvent, soit les étudiants forts de troisième année. La fréquence théorique de cette graphie est de 6,5%.

La graphie c dont la fréquence théorique est de 22,7% est, avec des proportions d'utilisation minimum et maximum de 20,0% et 33,5%, celle dont les fréquences empiriques se rapprochent le plus de la fréquence théorique.

Parmi les *autres* expressions graphémiques utilisées par les sujets de l'échantillon pour représenter le phonème /s/, certaines sont erronées. La plus courante, parmi les mauvaises graphies, est *ch*. Bien que cette expression graphémique contienne l'une des graphies qui peut représenter le phonème /s/, soit c, il serait étonnant que *ch* soit considéré, par les sujets de l'échantillon, comme l'une des expressions graphémiques pouvant être associée au phonème /s/. Il est plus probable que cette erreur provienne de la ressemblance phonétique de /s/ et /ʃ/. Parmi les mauvaises graphies, on a aussi retracé des cc, sc, x, cs. Il est intéressant de remarquer que chacune de ces graphies contient un graphème qui peut représenter le phonème /s/. Ces mauvaises graphies représentent, pour chacun des niveaux d'études et chacune des catégories de rendement en français, environ 1% des expressions graphémiques utilisées pour re-

présenter le phonème /s/. Le pourcentage d'erreurs étant si faible même en deuxième année, il semble que les enfants apprennent très tôt à distinguer les expressions graphémiques qui peuvent représenter ce phonème de celles qui ne le peuvent pas.

Pour l'ensemble des graphies du phonème /s/, on remarque que les groupes qui s'éloignent le plus des fréquences théoriques d'utilisation sont les étudiants forts et faibles de deuxième année. Par ailleurs, il se produit un progrès marqué dans le rapprochement des fréquences empiriques et théoriques entre la deuxième et la troisième année. Ce progrès est toutefois beaucoup plus marqué chez les forts que chez les faibles. En effet, la graphie ss par exemple, dont la fréquence théorique est de 32,0%, a été utilisée dans des proportions semblables par les étudiants forts et faibles de deuxième année, soit 6,7% et 5,0%. Par contre, tandis que cette graphie est déjà utilisée dans une proportion de 20,7% par les étudiants forts de troisième année, elle ne l'est encore que dans une proportion de 8,5% par les étudiants faibles du même niveau.

3.3.2.3 L'indice de popularité des graphies du phonème /ɛ/.

Il ressort de l'examen des pourcentages d'étudiants qui ont utilisé les diverses expressions graphémiques qui peuvent représenter le phonème /ɛ/ (tableau 3.10) que, tout comme on l'observait pour le phonème /s/, la presque totalité des étudiants a utilisé les graphies les plus fréquentes soit *e*, *ai* et *è*. Par contre, alors qu'un plus grand nombre d'étudiants parmi les plus jeunes ainsi que parmi les plus faibles de chaque niveau utilisaient, pour représenter le phonème /s/, les graphies auxquelles correspondent une faible fréquence théorique, on ne remarque pas de distribution semblable pour les graphies peu fréquentes du phonème /ɛ/. Par exemple, la graphie *ê* à laquelle ne correspond qu'une fréquence théorique de 6,5% a été utilisée par la majorité des forts de chaque niveau d'études alors qu'elle a été moins populaire parmi les faibles. En effet, c'est seulement en troisième et en sixième années que plus de la moitié des étudiants de cette catégorie ont eu recours à cette graphie pour représenter le phonème /ɛ/. Quant à la graphie *ei* qui ne représente elle aussi que rarement le phonème /ɛ/, au moins les trois quarts des étudiants de chaque niveau, à l'exception des faibles de deuxième année, l'ont utilisée.

A la suite de l'examen des données fournies par l'indice de popularité des graphies pour le phonème /s/, on avait supposé que ce qui peut distinguer les étudiants forts des

TABLEAU 3.10

POURCENTAGES D'ÉTUDIANTS DE CHAQUE NIVEAU D'ÉTUDES ET DE CHAQUE CATÉGORIE DE RENDEMENT EN FRANCAIS QUI ONT UTILISÉ LES DIVERSES EXPRESSIONS GRAPHÉMIQUES DU PHONÈME /ɛ/

NIVEAUX	RENDEMENTS	EXPRESSIONS GRAPHÉMIQUES					
		e %	*ai* %	*è* %	*ê* %	*ei* %	*autres* %
2	Forts	100	94	100	61	78	28
	Faibles	100	88	94	29	71	23
3	Forts	100	90	100	70	90	55
	Faibles	100	100	95	60	85	20
4	Forts	100	100	100	70	75	70
	Faibles	100	100	100	45	80	60
5	Forts	100	100	100	75	95	45
	Faibles	100	100	100	55	95	45
6	Forts	100	100	100	75	80	35
	Faibles	100	100	100	65	85	60
FRÉQUENCES THÉORIQUES		51,3	20,8	13,1	6,5		8,3

faibles est peut-être le fait que les forts auraient peu recours aux graphies rares justement parce qu'ils sauraient que ces graphies ne représentent que rarement un phonème donné. Or il ressort des données fournies par l'indice de popularité des graphies pour le phonème /c/ que plus d'étudiants forts que faibles ont utilisé au moins une fois dans l'épreuve orthographique, les graphies peu fréquentes de ce phonème. L'hypothèse énoncée lors de l'examen des données relatives au phonème /s/ serait donc infirmée. Il reste toutefois à analyser les informations contenues dans l'indice d'utilisation des graphies du phonème /c/ avant de se prononcer.

3.2.2.4 Indice d'utilisation des graphies du phonème /ɛ/.

Les proportions dans lesquelles les principales expressions graphémiques qui représentent le phonème /ɛ/ en position médiane ont été utilisées par les sujets de l'échantillon sont présentées au tableau 3.11. On constate, à l'examen de ces données, que la fréquence moyenne d'utilisation de la graphie *e* par l'ensemble des sujets de l'échantillon correspond à sa fréquence théorique en français, soit 51,3%. Le groupe qui y a le moins fréquemment recours, celui des étudiants faibles de quatrième année, l'utilise tout de même dans une proportion de 43,2%; celui qui l'utilise le plus souvent, c'est-à-dire le groupe des étudiants forts de sixième année, l'utilise dans une proportion de 55,1%.

La graphie *ai* est aussi utilisée dans une proportion qui se rapproche de sa fréquence théorique qui est de 20,8%. En effet, deux groupes seulement s'éloignent de façon marquée de cette proportion: ce sont les étudiants forts et faibles de deuxième année qui l'utilisent respectivement dans des proportions de 6,3% et 11,1%. En moyenne, cette graphie est utilisée dans une proportion de 16,9%.

Lors de la vérification de l'hypothèse E1, on avait remarqué que pour le phonème /ɛ/, non seulement le degré de rapprochement du code de l'ensemble des sujets de l'échantillon au code français est élevé mais il est, de plus, le plus

TABLEAU 3.11

POURCENTAGES D'UTILISATION DES EXPRESSIONS GRAPHÉMIQUES LES PLUS FRÉQUENTES POUR REPRÉSENTER LE PHONÈME /e/

NIVEAUX	RENDEMENTS	EXPRESSIONS GRAPHÉMIQUES					
		e %	*ai* %	*é* %	*ê* %	*autres bonnes* %	*autres mauvaises* %
2	Forts	51,1	6,3	34,1	2,7	2,7	3,1
	Faibles	47,4	11,1	28,5	0,6	2,7	9,7
3	Forts	51,3	16,9	20,5	3,3	5,3	2,7
	Faibles	53,7	17,0	18,4	2,7	3,3	4,9
4	Forts	51,9	15,9	21,7	3,5	4,5	2,5
	Faibles	43,2	26,6	21,4	1,2	4,7	2,9
5	Forts	52,5	17,6	20,0	3,0	4,4	2,5
	Faibles	51,6	23,2	17,1	2,6	3,9	1,6
6	Forts	55,1	15,9	17,7	3,6	3,5	4,2
	Faibles	55,0	18,8	19,4	2,8	2,0	2,0
Fréq. moyennes	Forts	52,4	14,5	22,8	3,2	4,1	3,0
	Faibles	50,2	19,3	21,0	2,0	3,3	4,2
Fréquences théoriques		51,3	20,8	13,1	6,5	8,3	0,0

élevé des trois phonèmes étudiés. Par ailleurs, il ressort de l'examen des données fournies par l'indice d'utilisation des graphies, que les pourcentages d'utilisation, par les sujets de l'échantillon, des expressions graphémiques *e* et *ai* se rapprochent de leur fréquence théorique. Or celles-ci totalisent plus de 70% des représentations du phonème /ɛ/ en français. Le fort degré de rapprochement du code des sujets de l'échantillon au code français pour ce phonème s'explique donc sans doute par le fait que le pourcentage d'utilisation de ces deux graphies qui sont les plus fréquemment utilisées en français pour représenter ce phonème, se rapproche de leur fréquence théorique.

On ne remarque toutefois pas un tel rapprochement pour les graphies plus rares. En effet, la graphie *è* à laquelle correspond une fréquence théorique de 13,1% est sur-utilisée par chacun des groupes. Avec une fréquence théorique de 6,5% et des fréquences d'utilisation minimum et maximum de 0,6% et 3,6%, la graphie *ê* est, quant à elle, sous-utilisée par chacun des groupes de l'échantillon.

Parmi les *autres* graphies qui ont été utilisées, certaines étaient erronées. La plus courante de ces graphies est *é*; on a noté aussi des *eai*, *ei*, *éi*, *èi*, *ès* et quelques graphies complètement étrangères au phonème /ɛ/ comme *an*, *eu*, *ar*. Ces dernières graphies ont été utilisées par les étudiants faibles de deuxième année. Dans ce groupe, envi-

ron 10% des expressions graphémiques utilisées pour représenter le phonème /ɛ/ étaient mauvaises. Dans les autres groupes, environ 3% des graphies étaient erronées.

On observait, lors de l'analyse de l'indice de similitude du code des sujets de l'échantillon au code français pour le phonème /ɛ/ (Hypothèse E1), un progrès marqué entre la deuxième et la troisième année. Or il ressort, de l'examen des valeurs prises par l'indice d'utilisation des graphies, que ce progrès est dû à un rapprochement sensible, entre la deuxième et la troisième année, des fréquences d'utilisation des graphies *ai*, *è* et les *autres* de leurs fréquences théoriques.

Lors de l'analyse de l'indice de popularité des graphies du phonème /ɛ/, on avait remarqué que plus d'étudiants forts que d'étudiants faibles ont utilisé au moins une fois, dans l'épreuve orthographique, les graphies qui représentent peu souvent le phonème /ɛ/. L'examen des données du tableau 3.11 sur l'indice d'utilisation des graphies révèle qu'au total, la graphie *è* à laquelle ne correspond qu'une fréquence théorique de 6,5% est plus utilisée par les forts que les faibles. Par contre, les *autres* graphies qui ont été ainsi regroupées parce qu'elles sont rares sont, au total, à peu près également utilisées par les forts et les faibles. Puisque les étudiants qui appartiennent à chacune des catégories de rendement en français ont des comportements diffé-

rents vis-à-vis de graphies qui ont en commun leur faible fréquence théorique, on en conclut que celle-ci ne semble pas être une variable qui puisse expliquer l'utilisation des diverses graphies rares par les étudiants forts et les étudiants faibles.

3.2.2.5 L'indice de popularité des graphies du phonème /o/

Il ressort de l'examen des expressions graphémiques utilisées par les sujets de l'échantillon pour représenter le phonème /o/ (tableau 3.12) que, comme pour les cas des phonèmes /s/ et /ɛ/, la presque totalité des individus a utilisé les graphies qui représentent le plus fréquemment ce phonème en français, soit *o* et *au*. Seuls les étudiants faibles de deuxième année ont fait exception à cette règle. En effet, 47% seulement des étudiants de ce groupe ont fait usage de la graphie *au* dans l'épreuve orthographique.

En ce qui concerne les expressions graphémiques plus rares, soit *eau*, *ô* et les *autres*, la distribution des pourcentages d'étudiants qui les ont utilisées varie de l'une à l'autre. A l'exception du groupe des faibles de deuxième année, la majorité des étudiants des autres groupes ont utilisé l'expression graphémique *eau* à laquelle correspond une fréquence théorique de 9,3%. Quant à la graphie *ô*, à laquelle correspond une fréquence théorique équivalente à celle de la graphie *eau*, soit 8,9%, les étudiants qui ont un rendement faible en français l'ont utilisée dans une proportion de 10% seulement, à l'exception toutefois du groupe des faibles de cinquième année qui y a eu recours dans une proportion de 45%. De la troisième à la cinquième année, environ 40% des étudiants qui ont un rendement supérieur ont utilisé la graphie *ô*.

TABLEAU 3.12

POURCENTAGES D'ÉTUDIANTS DE CHAQUE NIVEAU
ET DE CHAQUE CATÉGORIE DE RENDEMENT
EN FRANCAIS QUI ONT UTILISÉ
LES DIVERSES EXPRESSIONS GRAPHÉMIQUES
DU PHONEME /o/

NIVEAUX	RENDEMENTS	EXPRESSIONS GRAPHÉMIQUES				
		o %	*au* %	*eau* %	*ô* %	*autres* %
2	Forts	100	94	50	11	6
	Faibles	100	47	18	0	6
3	Forts	100	100	90	40	10
	Faibles	100	100	80	10	10
4	Forts	100	100	70	35	10
	Faibles	95	100	85	15	30
5	Forts	100	100	85	55	15
	Faibles	100	100	80	45	10
6	Forts	100	100	65	7	25
	Faibles	100	100	90	10	35
FRÉQUENCES THÉORIQUES		44,9	36,7	9,3	8,9	2,0

Cette variation dans le pourcentage d'étudiants forts et faibles qui ont utilisé des graphies dont la fréquence théorique est équivalente renforce ce qui a été observé dans le cas du phonème /ɛ/: la faible fréquence théorique de certaines expressions graphémiques ne pourrait expliquer l'utilisation différente des graphies rares par les étudiants forts et faibles.

3.2.2.6 Indice d'utilisation des graphies du phonème /o/

Les proportions dans lesquelles les principales expressions graphémiques qui représentent le phonème /o/ en position médiane ont été utilisées par les sujets de l'échantillon sont présentées au tableau 3.13. Il ressort, de l'examen de ces données, qu'avec une proportion moyenne d'utilisation de 46,5% et une fréquence empirique égale à 44,9%, la graphie *o* est utilisée dans des proportions qui sont très semblables à celle dans laquelle elle devrait être utilisée. Cette mesure est toutefois trompeuse puisque, d'une part, deux groupes d'étudiants, les forts et les faibles de deuxième année, l'utilisent dans des proportions aussi élevées que 79,2% et 90,1%. D'autre part, tous les autres groupes la sous-utilisent. Leurs fréquences d'utilisation qui varient de 31,5% à 39,8% ne s'éloignent toutefois pas tellement de la fréquence théorique de cette graphie.

L'expression graphémique *au* à laquelle correspond une fréquence théorique de 36,7% est très peu utilisée par les étudiants forts et faibles de deuxième année. Cette expression graphémique ne représente en effet que 14,0% et 4,3% des expressions graphémiques utilisées par ces groupes pour représenter le phonème /o/. Avec une fréquence d'utilisation qui se situe autour de 50%, chacun des autres groupes, par contre, sur-utilise cette graphie.

TABLEAU 3.13

POURCENTAGES D'UTILISATION DES EXPRESSIONS GRAPHÉMIQUES LES PLUS FRÉQUENTES POUR REPRÉSENTER LE PHONÈME /o/

NIVEAUX	RENDEMENTS	EXPRESSIONS GRAPHÉMIQUES					
		o %	*au* %	*eau* %	*ô* %	*autres bonnes* %	*autres mauvaises* %
2	Forts	79,2	14,0	4,8	0,0	0,2	1,8
	Faibles	90,1	4,3	1,7	0,0	0,3	3,6
3	Forts	37,9	48,7	11,3	0,8	0,2	1,1
	Faibles	37,2	51,6	8,0	0,2	0,2	2,8
4	Forts	39,1	48,7	8,2	2,0	0,2	1,8
	Faibles	38,4	48,6	8,2	0,3	1,1	3,4
5	Forts	37,2	50,0	8,6	3,3	0,3	0,6
	Faibles	34,6	53,1	8,9	1,3	0,2	1,9
6	Forts	39,8	50,3	6,8	1,3	0,6	1,2
	Faibles	31,5	52,5	9,4	2,3	0,8	3,5
Fréq. moyennes	Forts	46,6	42,3	7,9	1,5	0,3	1,3
	Faibles	46,3	49,4	7,2	0,8	0,5	3,0
Fréquences théoriques		44,9	36,7	9,3	8,9	2,0	0,0

Comme pour le cas de la graphie *au*, les étudiants forts et faibles de deuxième année sous-utilisent l'expression graphémique *eau*. Ces étudiants l'utilisent en effet dans des proportions de 4,8% et 1,7% alors qu'elle a une fréquence théorique de 9,3%. Quant aux autres groupes, ils respectent assez bien sa fréquence théorique puisqu'ils y ont recours dans des proportions qui varient de 6,8% à 11,3%.

Enfin, la graphie *ô* est nettement sous-utilisée par chacun des groupes d'étudiants. La fréquence d'utilisation moyenne de cette graphie est en effet de 1,1% alors que sa fréquence théorique est de 8,9%.

Parmi les mauvaises graphies auxquelles les sujets de l'échantillon ont eu recours, les plus souvent utilisées sont *ou* et *on* dans chacune desquelles on retrouve une graphie du phonème /o/. Dans chaque groupe d'etudiants, environ 2% des expressions graphémiques utilisées pour représenter le phonème /o/ étaient erronées.

On se rappelle que l'indice de similitude au code français pour le phonème /o/, qui a été établi lors de la vérification de l'hypothèse E1, affichait une différence importante entre la deuxième et la troisième année et une relative stabilisation de la troisième à la sixième année. Il ressort de l'examen des données fournies par l'indice d'utilisation des graphies de ce phonème que ce saut de

la deuxième à la troisième année, suivi d'une stabilisation, peut être observé, au tableau 3.13, pour chacune des expressions graphémiques du phonème /o/.

3.3.3 Conclusion

Selon l'hypothèse E2, l'individu qui fait l'apprentissage du français écrit apprendrait progressivement à associer, à chaque phonème, les diverses expressions graphémiques qui peuvent le représenter. Pour vérifier cette hypothèse, trois outils d'analyse complémentaires ont été développés. Ces outils sont l'indice d'étendue du code, l'indice de popularité des graphies et l'indice d'utilisation des graphies.

L'examen de l'indice d'étendue du code pour chacun des phonèmes étudiés révèle que l'habileté à associer, à chaque phonème, les diverses expressions graphémiques qui peuvent le représenter serait acquise en général très tôt, vraisemblablement au cours de la première année d'études. On remarque en effet que l'étendue du code pour un phonème donné est, à partir de la deuxième année, sensiblement la même pour tous les étudiants, quels que soient le niveau d'études et la catégorie de rendement en français. Seuls les étudiants forts et faibles de deuxième année possèdent, pour le phonème /o/, un indice d'étendue plus faible que celui des étudiants des niveaux supérieurs.

Pour étudier plus finement la progression de l'étendue du code phono-graphique d'un élève, il y aurait donc lieu de choisir des sujets plus jeunes et de les soumettre à l'épreu-

ve écrite plus d'une fois au cours de l'année. Que l'on retienne cependant que dès la deuxième année, les étudiants forts et faibles savent, en général, associer, à chaque phonème, les diverses expressions graphémiques qui peuvent le représenter.

Lors de la vérification de l'hypothèse E1, on constatait que les valeurs prises par l'indice de similitude indiquaient que le code des sujets de l'échantillon se rapproche du code français de façon marquée entre la deuxième et la troisième année et de façon plus lente au-delà de la troisième année. Les valeurs prises par cet indice indiquaient de plus que pour chaque phonème, le code des étudiants forts de chacun des niveaux d'études se rapproche davantage du code français que celui des étudiants faibles.

Or, les valeurs prises par l'indice d'étendue du code révèlent que la maîtrise des correspondances phonèmes - graphèmes est généralement acquise, chez les forts comme chez les faibles, dès la deuxième année. L'évolution du code des sujets de l'échantillon vers le code français qui a été observée lors de la vérification de l'hypothèse E1 ne peut donc pas être expliquée par une progression dans la maîtrise des correspondances phonèmes - graphèmes.

L'indice d'utilisation des graphies peut cependant fournir des informations sur l'indice de similitude. En ef-

fet, alors que l'indice de similitude constitue une synthèse de l'évolution du code phono-graphique pour les phonèmes /s/, /k/ et /o/, l'indice d'utilisation des graphies permet d'analyser ce phénomène puisqu'il présente les pourcentages d'utilisation de chacune des expressions graphémiques des phonèmes étudiés. La convergence des fréquences d'utilisation et des fréquences théoriques telle que démontrée par l'indice de similitude pourrait provenir de la convergence d'un nombre plus ou moins grand de graphies. Il est possible en effet que le rapprochement observé s'explique par la convergence asymptotique d'une ou de deux graphies comme il est possible aussi que cette convergence soit le produit d'un rapprochement de chacune des graphies.

Un examen des données de l'indice d'utilisation des graphies indique qu'en général, pour les graphies les plus fréquemment utilisées en français, on observe un rapprochement important des fréquences empirique et théorique dès les premières années du cours primaire, surtout chez les étudiants forts. Quant aux graphies plus rares, leur convergence ne suit pas un modèle unique.

Le saut observé dans les valeurs de l'indice de similitude entre la deuxième et la troisième année, par exemple, serait attribuable, pour le phonème /s/, au rapprochement entre les fréquences d'utilisation et les fréquences théoriques, chez les forts et les faibles, de la graphie s qui est

une graphie fréquente en français et de la graphie ç qui est moins courante. Chez les forts, on observe également une convergence de la graphie *aa* à laquelle est rattachée une fréquence théorique élevée.

Pour le phonème /ɛ/, les rapprochements les plus marqués entre les fréquences d'utilisation et les fréquences théoriques sont observées pour les graphies *ai* et *è* qui sont plutôt fréquentes en français et les graphies plus rares qui ont été regroupées sous la rubrique *autres*. Pour le phonème /o/, ce sont, de la deuxième à la troisième année, pour les graphies à fréquence élevée *o* et *au* et la graphie plus rare *eau* que l'on remarque un rapprochement sensible entre les fréquences d'utilisation et les fréquences théoriques.

Ainsi la convergence de la plupart des graphies les plus fréquentes et celle de quelques graphies plus rares expliquent l'évolution des valeurs prises par l'indice de similitude entre la deuxième et la troisième année. L'examen des valeurs prises par l'indice d'utilisation des graphies pourrait permettre d'expliquer également d'autres particularités de l'évolution de l'indice de similitude comme, par exemple, la convergence plus grande du code des étudiants forts ou encore la stabilisation ou la régression observées entre la cinquième et la sixième année.

L'indice d'utilisation des graphies permet aussi de constater que les enfants qui apprennent à orthographier le français parviennent très tôt à distinguer, parmi toutes les expressions graphémiques, celles qui peuvent représenter un phonème donné de celles qui ne peuvent pas. A l'exception du phonème /ɛ/ pour lequel environ 7% des expressions graphémiques utilisées par les étudiants de deuxième année sont erronées, pour les niveaux supérieurs et pour les autres phonèmes, le pourcentage d'erreur est inférieur à 2,5%. Une analyse sommaire des erreurs indique que les graphies choisies représentent, le plus souvent, des phonèmes qui se rapprochent, phonétiquement, du phonème étudié ou font appel à des lettres qui font partie d'expressions graphémiques qui sont utilisées pour la représentation du phonème.

Quant aux valeurs prises par l'indice de popularité des graphies, elles ont permis de se rendre compte que certaines graphies correctes mais moins courantes sont utilisées différemment par les forts et les faibles. Par exemple, 30% des forts utilisent la graphie *ô* alors que seulement 16% des faibles en font autant. Quant à la graphie *oc*, 17% des forts s'en servent alors que cette proportion s'élève à 31% dans le cas des faibles. Il semble donc que les étudiants forts n'aient pas toujours mieux intégré que les étudiants faibles, les fréquences théoriques des graphies plus rares. De telles observations peuvent inciter à penser que ce qui distinguerait les étudiants dont le rendement en fran-

çais est supérieur pourrait être autre chose que l'absorption, consciente ou non, des fréquences théoriques des expressions graphémiques.

L'examen des résultats obtenus de la vérification de l'hypothèse E2 ne permet pas de confirmer cette hypothèse. Les données indiquent en effet qu'il n'y a généralement pas de progrès entre la deuxième et la sixième année dans la maîtrise des associations phonèmes-expressions graphémiques. Cependant, on constate que dès la deuxième année, les principales associations sont déjà maîtrisées de sorte qu'on ne peut s'attendre à observer un progrès marqué entre la deuxième et la sixième année. Si l'expérimentation avait été faite auprès de sujets plus jeunes, l'hypothèse E2 aurait alors peut-être pu être confirmée.

On a vu, lors de la vérification de l'hypothèse E1, que le rapprochement entre les fréquences d'utilisation et les fréquences théoriques ne peut être attribuable à la seule mémorisation de la représentation graphémique des mots de la langue. La vérification de l'hypothèse E2 indique que le rapprochement ne peut, non plus, être dû à une progression dans la maîtrise des correspondances phonème -graphème puisque celles-ci sont acquises dès la deuxième année. Pour expliquer ce rapprochement, on peut explorer la possibilité qu'il soit dû à une maîtrise progressive des règles de formation des suites graphémiques ou encore, à une intégration, consciente ou non, des fréquences

théoriques des diverses graphies qui peuvent représenter un phonème.

Dans la section qui suit, on tentera d'établir le rôle de la maîtrise des règles de formation des suites graphémiques dans le rapprochement entre les fréquences d'utilisation des diverses graphies et leur fréquence théorique correspondante.

3.4 Vérification de l'hypothèse E3

HYPOTHÈSE E3 : L'individu qui apprend à orthographier le français acquiert, au moins inconsciemment, la maîtrise progressive des règles de formation des suites graphémiques correctes dans cette langue.

Les règles de formation de suites graphémiques telles qu'on les a définies sont des formules qui indiquent quelles graphies doivent être utilisées pour représenter un phonème dans un environnement donné. Ce concept de règle diffère de deux façons de celui généralement utilisé en linguistique. D'abord, en linguistique générative transformationnelle, les règles sont déterministes alors que dans cette étude, elles ne font que restreindre l'ensemble des solutions possibles. Deuxièmement, les règles employées dans ce travail ne sont pas ordonnées. Ces deux différences impliquent qu'à partir d'une suite phonologique, on peut produire plusieurs suites graphémiques légitimes. De plus, les règles formulées ne proposent pas un modèle intégré. Elles ont été sélectionnées de façon à constituer un échantillon qui doit servir à vérifier l'hypothèse E3. Ainsi, pour vérifier cette hypothèse, on a formulé des règles de formation de suites graphémiques correctes qui contiennent des représentations des phonèmes /s/ et /ɛ/.[1] Ces règles tiennent compte uniquement des environnements gra-

(1) Les règles qui ont été formulées pour le phonème /ɔ/ comptaient chacune un nombre d'occurrences trop restreint pour être retenues.

phémiques immédiats de ces phonèmes, c'est-à-dire du graphème qui précède ou qui suit immédiatement l'expression graphémique qui représente le phonème étudié.

Chaque environnement peut être constitué de voyelles ou de consonnes. On a réuni, sous l'appellation "voyelle" (V), les éléments suivants, qu'ils soient ou non surmontés d'un accent:

$$V = \{ a, e, i, o, u, y \}$$

On a réuni, sous l'appellation "consonne" (C), l'ensemble des autres éléments de l'alphabet:

$$C = \{ b, c, d, f, g, h, j, k, l, m, n, p, q, r, s, t, v, w, x, z. \}$$

Pour décrire l'environnement graphémique des phonèmes étudiés, on fait parfois référence à la notion de "syllabe". Pour diviser les logatomes en syllabes, on a appliqué les trois règles de Grevisse:

"1° Une consonne placée entre deux voyelles introduit une nouvelle syllabe: *Cha-ri-ot, la-mi-noir...*

2° De deux consonnes placées entre deux voyelles, la première appartient à la syllabe précédente, la seconde, à la syllabe suivante: *ar-gent, mas-sif, ar-bus-te.*

Toutefois, *l* ou *r*, précédés d'une consonne autre que *l* ou *r*, forment avec cette consonne qui les précède, un groupe insé-

parable qui commence une syllabe: *li-bre, no-ble, en-tre-pren-dre.* Les groupes *ch, ph, gn, th* sont inséparables: *or-tho-gra-phe...*

3e Quand il y a trois consonnes consécutives à l'intérieur d'un mot, ordinairement les deux premières terminent une syllabe, l'autre commence une nouvelle syllabe. Toutefois les groupes *bl, cl, fl*, etc... commencent ordinairement une syllabe: *obs-ti-né, comp-ter, es-ca-lan-dre, ar-bris-seau."* (Grevisse, 1975, p.53)

Les règles qui ont été formulées pour vérifier l'hypothèse E3 sont parfois larges en ce sens qu'elles réfèrent, chacune, à plusieurs environnements différents. Il aurait été possible de décomposer ces règles en sous-règles qui auraient traité les environnements séparément. Ces sous-règles auraient présenté l'avantage d'être plus précises. Cependant, dans l'épreuve orthographique utilisée pour vérifier les hypothèses de travail, certains environnements apparaissent peu souvent. Conséquemment, pour permettre un traitement statistique fiable, il s'est avéré nécessaire de formuler des règles de façon à ce que le nombre d'occurrences de chacune soit suffisamment élevé. Si les résultats portant sur les règles retenues ne permettent pas d'infirmer l'hypothèse E3, on pourrait, pour mieux s'assurer de la justesse des conclusions, mettre à nouveau cette hypothèse à l'épreuve mais, cette fois, au moyen de règles plus nombreuses et plus raffinées. Par contre, si une analyse des données à partir des règles formulées infirme l'hypothèse à vérifier,

une analyse à partir de règles plus raffinées conduirait, à fortiori, à la même conclusion.

L'épreuve orthographique ayant été construite de façon à respecter les fréquences, dans la langue française, des environnements phonologiques des phonèmes étudiés, le nombre d'occurrences de chacune des règles ne peut, par conséquent, être le même. Il varie en effet en fonction du nombre d'occurrences des environnements phonologiques retenus pour la construction de l'épreuve orthographique. De même, le nombre d'occurrences d'une règle donnée peut varier d'un individu à l'autre. Le choix d'une expression graphémique pour représenter un phonème est en effet fonction de l'environnement graphémique de ce phonème. Ainsi, si, ayant à représenter la suite phonologique /asɛ/ par exemple, un individu décide de choisir la graphie *ai* pour représenter le phonème /ɛ/ alors qu'un autre choisit la graphie *e*, chacun ne se référera pas à la même règle pour représenter le phonème /s/. C'est pourquoi, lorsqu'on parle du nombre d'occurrences de chacune des règles, on parle de nombre moyen par étudiant.

Dans cette section, on présentera le taux de réussite de chacune des règles analysé en fonction des variables indépendantes *niveau d'études* et *rendement en français*.

3.4.1 Maîtrise de la règle R1

$$R1: \quad /s/ \dashrightarrow \begin{Bmatrix} ss \\ ç \end{Bmatrix} \Big/ V __ \begin{Bmatrix} a \\ o \\ u \end{Bmatrix}$$

Règle R1 : Le phonème /s/ doit être représenté par l'expression graphémique ç ou ss s'il est précédé d'une voyelle et suivi de la voyelle *a*, *o* ou *u*.

Exemples : *glissade, chaussor dessus, deça, hameçon, déçu.*

Cette règle compte, en moyenne, dix occurrences par étudiant. Son taux de réussite, présenté en fonction des variables indépendantes *niveau d'études* et *rendement en français*, apparaît au tableau 3.14. Si on considère d'abord les résultats en fonction de la variable *niveau d'études*, on remarque que le plus grand saut dans l'apprentissage de cette règle se produit entre la deuxième et la troisième année d'études, le taux de réussite passant alors de 17,6% à 51,7%. Au-delà de la troisième année, il y a ou régression, ou apprentissage très lent, si bien que les individus de sixième année n'ont en moyenne réussi cette règle qu'à 57,1%. On peut donc dire qu'à la fin du cours primaire, cette règle n'est pas maîtrisée par l'ensemble des sujets de l'échantillon.

Un examen des résultats présentés en fonction de la variable *rendement en français* montre qu'il y a peu de différence dans le taux de réussite de cette règle chez les forts

TABLEAU 3.14

TAUX MOYENS DE RÉUSSITE À LA REGLE R1
EN FONCTION DU NIVEAU D'ÉTUDES ET
DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS FORTS %	FAIBLES %	MOYENNE %
2	18,6	16,5	17,6
3	63,6	39,8	51,7
4	64,8	30,6	47,7
5	62,4	35,4	48,9
6	57,5	56,7	57,1
Moyenne	53,4	35,8	44,6

et les faibles de deuxième année et on peut dire qu'avec des moyennes de 18,6% et de 16,5%, la règle R1 n'est pas du tout maîtrisée à ce niveau. Entre la deuxième et la troisième année, il se produit un progrès important pour les deux groupes mais ce progrès est beaucoup plus grand pour le groupe des forts que pour celui des faibles, le taux de réussite passant de 18,6% à 63,6% chez les forts et de 16,5% à 39,8% chez les faibles. De la quatrième à la sixième année, on observe une légère régression dans le taux de réussite de cette règle chez les forts. Chez les faibles, on observe aussi une régression entre la troisième et la cinquième année. Contrairement à ce qui se produit chez les forts, on observe, par contre, entre la cinquième et la sixième année, un progrès marqué chez les faibles, le taux de réussite passant alors de 35,4% à 56,7%. A cause de ce progrès, les forts et les faibles de sixième année obtiennent finalement un taux de réussite équivalent, soit 57,5% pour les forts et 56,7% pour les faibles. Il est intéressant de remarquer que les étudiants forts de troisième année, avec un taux de réussite de 63,6%, maîtrisent davantage la règle R1 que les étudiants faibles de sixième année dont le taux de réussite est de 56,7%.

Il ressort des résultats de l'analyse de variance effectuée à partir du taux de réussite à la règle R1 et qui sont présentés au tableau 3.15 que l'interaction *rendement en français - niveau d'études* est significative. Ce résultat reflète l'écart inégal entre le taux de réussite des forts

TABLEAU 3.15

RÉSULTATS DE L'ANALYSE DE VARIANCE
EFFECTUÉE À PARTIR DES RÉSULTATS OBTENUS
À LA RÈGLE R1

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	14783,0	4	22,3	0,001
Rendement en français	17887,1	1	26,9	0,001
Interaction niv.-rend.	1965,7	4	2,9	0,021
Résidus	663,7	183		

et des faibles d'un niveau à l'autre. Les taux de réussite de ces deux groupes ont été illustrés au graphique 3.01 afin de mieux faire ressortir les différences dans l'évolution du taux de réussite d'un niveau d'études à l'autre.

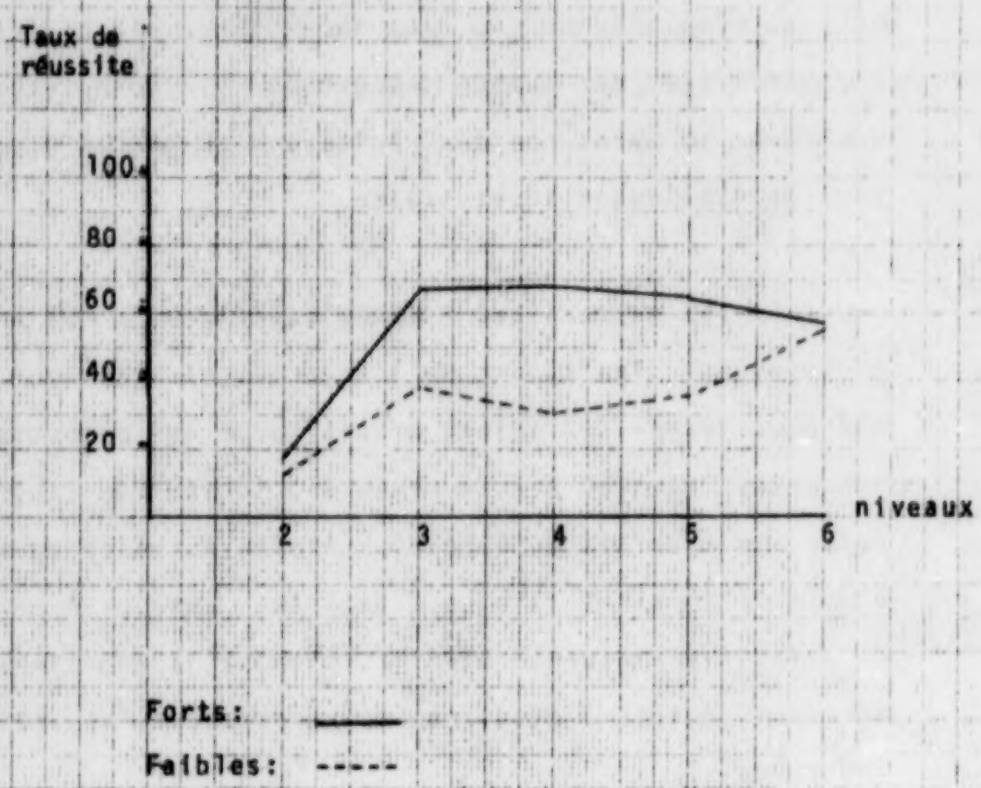
La réussite de la règle R1 exige que le phonème /s/ soit représenté par les expressions graphémiques ss ou ç. Or l'étude qui a été faite, lors de la vérification de l'hypothèse E2 de l'indice d'utilisation des graphies du phonème /s/, révélait que ces deux expressions graphémiques sont sous-utilisées par chacun des groupes d'étudiants. Conséquemment, au moins une des règles qui exige ces graphies sera partiellement insatisfaite.

Si les expressions graphémiques ss et ç qui ont été utilisées par les sujets de l'échantillon l'avaient toujours été pour former des suites graphémiques qui respectent les exigences imposées par la règle R1, le taux de réussite à cette règle aurait dû être plus élevé, au moins pour les étudiants forts et faibles de sixième année. A ce niveau, en effet, la fréquence d'utilisation de la graphie ss par les deux groupes d'étudiants se rapproche de sa fréquence théorique.

Le faible taux de réussite, même pour les étudiants de sixième année, peut s'expliquer d'au moins deux façons. Premièrement, l'utilisation des graphies exigées pour la réussi-

GRAPHIQUE 3.01

INTERACTION ENTRE LE NIVEAU D'ÉTUDES ET LE RENDEMENT EN FRANÇAIS POUR LA RÈGLE R1



te de la règle R1 peut avoir été faite, au moins dans certains cas, dans des contextes dans lesquels les étudiants ont cru, à tort, qu'elles étaient exigées. Deuxièmement, il est possible que ce que l'étudiant a progressivement intégré, consciemment ou non, ne soit pas la maîtrise des règles mais la fréquence théorique des diverses expressions graphémiques. Dans ce cas, une fréquence d'utilisation des graphies qui se rapprocherait de la fréquence théorique n'entraînerait un haut taux de réussite à une règle donnée que par l'effet du hasard surtout si la somme des fréquences théoriques des graphies exigées pour la réussite de cette règle est faible.

3.4.2 Maîtrise de la règle R2

$$R2: \quad /s/ \dashrightarrow \begin{Bmatrix} c \\ ss \\ t \end{Bmatrix} \Big/ V __ \begin{Bmatrix} e \\ i \\ y \end{Bmatrix}$$

Règle R2 : Le phonème /s/ doit être représenté par l'expression graphémique *c*, *ss* ou *t* quand il est précédé d'une voyelle et suivi de la voyelle *e*, *i* ou *y*.

Exemples : *décembre*, *décider*, *bicyclette*, *assez*, *mission*, *solution*.

Cette règle compte, en moyenne, trente-huit occurrences par étudiant. Son taux de réussite, présenté en fonction des variables indépendantes *niveau d'études* et *rendement en français*, apparaît au tableau 3.16. On remarque d'abord, à l'examen des résultats présentés en fonction de la variable *niveau d'études* qu'alors qu'il y a un progrès important entre la deuxième et la troisième année, le taux de réussite passant de 42,1% à 58,9%, le niveau de maîtrise de cette règle est assez stable de la quatrième à la sixième année, le taux de réussite, pour la quatrième année étant de 54,4% et pour la sixième année, de 54,8%.

Si on examine maintenant les résultats présentés en fonction de la variable *rendement en français*, on observe que les forts et les faibles de deuxième année ont un taux de réussite comparable. Le progrès qui se situe entre la deuxième et la troisième année est toutefois beaucoup plus important chez

TABLEAU 3.16

TAUX MOYENS DE RÉUSSITE À LA RÈGLE R2
EN FONCTION DU NIVEAU D'ÉTUDES ET
DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS %	FAIBLES %	MOYENNE %
2	41,9	42,3	42,1
3	67,2	50,6	58,9
4	63,4	45,4	54,4
5	57,7	48,7	53,2
6	52,0	57,6	54,8
Moyenne	56,4	48,9	52,7

les forts que chez les faibles, le taux de réussite passant de 41,9% à 67,2% chez les forts et de 42,3% à 50,6% chez les faibles. Alors que chez les forts, le degré de maîtrise de la règle R2 régresse de la troisième à la sixième année, on remarque, chez les faibles, un progrès appréciable entre la cinquième et la sixième année si bien que le taux de réussite des étudiants faibles de sixième année qui est de 57,6% est finalement supérieur à celui des forts du même niveau dont le taux de réussite n'est que de 52,0%. Avec un taux de réussite de 67,2%, il est intéressant de remarquer que les étudiants forts de troisième année obtiennent le taux de réussite le plus élevé de tous les groupes. Ce taux n'est toutefois pas très élevé et on peut dire qu'en général, la règle R2 n'est pas bien maîtrisée par les groupes d'étudiants qui ont été soumis à l'épreuve orthographique.

Il ressort de l'analyse de variance effectuée à partir du taux de réussite de cette règle (Tableau 3.17) que la variable *niveau d'études* n'a pas d'effet significatif sur le taux de réussite. Les résultats de cette analyse de variance confirment toutefois l'effet de la variable *rendement en français*. Bien que l'écart entre le taux de réussite des forts et des faibles ne soit pas constant d'un niveau d'études à l'autre, l'interaction *niveau d'études - rendement en français* n'est pas significative.

TABLEAU 3.17

RÉSULTATS DE L'ANALYSE DE VARIANCE
EFFECTUÉE À PARTIR DES RÉSULTATS OBTENUS
À LA RÈGLE R2

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	1517,9	4	2,23	0,068
Rendement en français	4554,7	1	6,68	0,011
Interaction niv.-rend.	1024,1	4	1,50	0,204
Résidus	682,2	183		

La réussite de la règle R2 exige que le phonème /s/ soit représenté par les expressions graphémiques c, ss ou t. Comme pour le cas de la règle R1, on aurait pu s'attendre à ce que le taux de réussite à la règle R2 soit plus élevé, au moins pour les étudiants de sixième année. Leur fréquence d'utilisation des graphies ss et c est en effet très près de leur fréquence théorique correspondante. Quant à la graphie t qui à cause de sa faible fréquence théorique, a été classée dans le groupe des *autres* graphies pour le calcul de l'indice d'utilisation, on ne sait pas si elle a été sur- ou sous-utilisée. On sait cependant que les *autres* graphies représentent une faible fréquence empirique pour les étudiants de ce niveau. L'importance de l'utilisation de t sur le taux de réussite à la règle R2 est donc minime.

Ainsi, compte tenu des fréquences d'utilisation des graphies c, ss et t, on aurait pu s'attendre à ce que le taux de réussite à la règle R2 soit plus élevé, au moins pour les étudiants de sixième année. Les raisons de cet écart sont les mêmes que celles qui ont été énoncées dans le cas de la règle R1.

3.4.3 Maîtrise de la règle R3

$$R3: \quad /s/ \dashrightarrow \left\{\begin{matrix} c \\ ś \end{matrix}\right\} \Big/ C^1 __ \left\{\begin{matrix} e \\ i \\ y \end{matrix}\right\} \quad \text{où } C^1 \neq ś$$

Règle R3 : Le phonème /s/ doit être représenté par l'expression graphémique c ou ś quand il est précédé d'une consonne différente de ś et suivi de la voyelle *e*, *i* ou *y*.

Exemples : *pincer*, *principal*, *encyclique*, *conserver*, *insigne*.

Cette règle compte, en moyenne, treize occurrences par étudiant. Son taux de réussite, présenté en fonction des variables indépendantes *niveau d'études* et *rendement en français*, apparaît au tableau 3.18.

Il ressort de ces résultats que la règle R3 est maîtrisée dès la deuxième année, le taux de réussite étant de 93,4%. De plus, les étudiants considérés comme faibles en français la maîtrisent presque aussi bien que ceux dont le rendement dans cette matière est supérieur.

Les résultats de l'analyse de variance présentés au tableau 3.19 confirment l'absence de relation entre la variable indépendante *niveau d'études* et le taux de réussite, la règle étant maîtrisée dès la deuxième année. L'effet de la

TABLEAU 3.18

TAUX MOYENS DE RÉUSSITE À LA RÈGLE R3
EN FONCTION DU NIVEAU D'ÉTUDES ET
DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS %	FAIBLES %	MOYENNE %
2	95,3	91,5	93,4
3	94,3	93,9	94,1
4	93,3	91,2	92,3
5	97,9	91,9	94,9
6	92,4	87,5	90,0
Moyenne	94,6	91,2	92,9

TABLEAU 3.19

RÉSULTATS DE L'ANALYSE DE VARIANCE
EFFECTUÉE À PARTIR DES RÉSULTATS OBTENUS
À LA RÈGLE R3

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	148,9	4	1,01	0,404
Rendement en français	574,4	1	3,89	0,050
Interaction niv.-rend.	48,9	4	0,33	0,857
Résidus	147,5	183		

variable *rendement en français* est toutefois significatif.

L'une des graphies exigées pour la réussite de la règle R3 est la graphie s. Or comme l'indique l'indice d'utilisation des graphies, cette expression graphémique a été largement sur-utilisée dans l'épreuve orthographique, particulièrement par les étudiants de deuxième année qui y ont eu recours deux fois plus souvent que ne l'exigeait la langue française standard. En fait, les deux graphies exigées pour la réussite de la règle R3, soit s et c, représentent près de 90% des graphies utilisées par les étudiants de deuxième année pour représenter les diverses occurrences du phonème /s/. Par conséquent, même si ces étudiants ne connaissent pas la règle R3, le seul hasard suffirait à expliquer qu'ils aient obtenu un score très élevé à cette règle. Il est donc difficile de savoir si la règle R3 est véritablement maitrisée dès la deuxième année ou si c'est la somme des fréquences d'utilisation des graphies qu'elle exige qui biaise les résultats obtenus.

3.4.4 Maîtrise de la règle R4

$$R4: \quad /s/ \longrightarrow \left\{ \begin{matrix} ç \\ ś \end{matrix} \right\} / \; C^1 __ \left\{ \begin{matrix} a \\ o \\ u \end{matrix} \right\} \quad \text{où } C^1 \neq ś$$

Règle R4 : Le phonème /s/ doit être représenté par l'expression graphémique ç ou ś s'il est précédé d'une consonne différente de ś et suivi de la voyelle *a*, *o* ou *u*.

Exemples : *français*, *balançoire*, *perçu*, *insatisfait*, *chanson*, *ensuite*.

Cette règle compte, en moyenne, sept occurrences par étudiant. Son taux de réussite, présenté en fonction des variables indépendantes *niveau d'études* et *rendement en français*, apparaît au tableau 3.20. L'apprentissage de la règle R4 semble être maîtrisé dès le niveau de la deuxième année puisque le taux moyen de réussite pour ce niveau est de 88,8%. Les étudiants dont le rendement en français est supérieur ont un taux de réussite plus élevé que les étudiants dont le rendement est plus faible en deuxième, en cinquième et en sixième années. L'effet de la variable *rendement en français* sur le taux de réussite n'est toutefois pas significatif comme le montre le tableau 3.21 qui présente les résultats d'une analyse de variance effectuée à partir du taux de réussite à la règle R4. La règle R4 étant maîtrisée dès la deuxième année, la variable *niveau d'études* n'est pas, non plus, significative.

TABLEAU 3.20

TAUX MOYENS DE RÉUSSITE À LA RÈGLE R4
EN FONCTION DU NIVEAU D'ÉTUDES ET
DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS %	FAIBLES %	MOYENNE %
2	91,7	85,9	88,8
3	92,8	91,6	92,2
4	85,6	88,4	87,0
5	95,4	84,6	90,0
6	93,1	86,3	89,7
Moyenne	91,7	87,4	89,6

TABLEAU 3.21

RÉSULTATS DE L'ANALYSE DE VARIANCE
EFFECTUÉE À PARTIR DES RÉSULTATS OBTENUS
À LA REGLE R4

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	551,0	4	1,41	0,232
Rendement en français	984,2	1	2,52	0,114
Interaction niv.-rend.	242,2	4	0,62	0,649
Résidus	390,5	183		

Comme pour la règle R3, la réussite de la règle R4 exige l'utilisation de la graphie à qui a été largement surutilisée par tous les groupes d'étudiants, particulièrement par ceux de deuxième année qui y ont eu recours dans une proportion d'environ 65%. Comme pour la règle R3, quoique dans une moindre mesure, il est donc difficile de savoir si cette règle est véritablement maîtrisée dès la deuxième année.

3.4.5 Maîtrise de la règle R5

$$R5: \quad /\varepsilon/ \dashrightarrow \begin{Bmatrix} e \\ ai \end{Bmatrix} \; / C __ C+ \quad (1)$$

Règle R5 : Le phonème /ɛ/ doit être représenté par l'expression graphémique *e* ou *ai* s'il est précédé d'une consonne et suivi d'une consonne qui termine la syllabe à laquelle il appartient.

Exemples: *sel, sec, nef, pair, clair.*

Cette règle comporte au moins une exception puisque le phonème /ɛ/ ne peut être représenté par l'expression graphémique *ai* quand il est suivi de la consonne *l* qui termine la syllabe à laquelle il appartient. Exemple: *portail* qui se lit /portaj/ et non /portɛj/. La règle R5 a donc été divisée en deux sous-règles qui peuvent être exprimées de la façon suivante:

$$R5.1: \quad /\varepsilon/ \dashrightarrow \begin{Bmatrix} e \\ ai \end{Bmatrix} / C __ C^1+ \quad \text{où } C^1 \neq l$$

$$R5.2: \quad /\varepsilon/ \dashrightarrow e \; / C __ l+$$

La sous-règle R5.1 compte, en moyenne, trente-huit occurrences par étudiant. La sous-règle R5.2 compte, en moyenne, une occurrence par étudiant. A cause de ce nombre restreint d'occurrences pour la sous-règle R5.2, les deux sous-règles ont été analysées ensemble sous le nom de règle R5. Le taux moyen de réussite de cette règle présenté en fonction des

---

(1) Le signe + indique la fin d'une syllabe.

variables indépendantes *niveau d'études* et *rendement en français* apparaît au tableau 3.22. Il ressort des résultats de ce tableau que, mis à part le recul que l'on observe entre la troisième et la quatrième année alors que le taux de réussite passe de 72,6% à 69,3%, l'apprentissage est progressif de la deuxième à la sixième année alors que le taux passe de 71,0% à 84,4%. L'écart entre les résultats des forts et des faibles est assez important en deuxième, troisième et quatrième années. Les faibles réussissent toutefois à rejoindre les forts en cinquième année alors que les taux de réussite sont équivalents, soit 81,4% pour les forts et 81,0% pour les faibles. En sixième année, les forts reprennent une légère avance sur les faibles. Il est intéressant de remarquer que les étudiants forts de deuxième année maîtrisent mieux la règle R5 que les étudiants faibles de quatrième année.

Les résultats de l'analyse de variance présentés au tableau 3.23 indique que la variable *niveau d'études* a un effet significatif sur le taux de réussite de cette règle. Toutefois, la variable *rendement en français* de même que l'interaction *niveau d'études - rendement en français* n'ont pas d'effet significatif.

TABLEAU 3.22

TAUX MOYENS DE RÉUSSITE À LA RÈGLE R5
EN FONCTION DU NIVEAU D'ÉTUDES ET
DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS %	FAIBLES %	MOYENNE %
2	75,9	66,1	71,0
3	76,3	68,8	72,6
4	73,7	64,8	69,3
5	81,4	81,0	81,2
6	86,7	82,1	84,4
Moyenne	78,8	72,6	75,7

TABLEAU 3.23

RÉSULTATS DE L'ANALYSE DE VARIANCE
EFFECTUÉE À PARTIR DES RÉSULTATS OBTENUS
À LA RÈGLE R5

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	1999,2	4	3,12	0,017
Rendement en français	719,6	1	1,12	0,292
Interaction niv.-rend.	209,7	4	0,33	0,860
Résidus	643,5	183		

3.4.6 Maîtrise de la règle R6

$$R6: \quad /\varepsilon/ \dashrightarrow \left\{ \begin{matrix} \grave{e} \\ \grave{e} \\ ai \\ aî \\ ei \end{matrix} \right\} / \ C __ \ \# \ C$$

Règle R6 : Le phonème /ɛ/ doit être représenté par l'expression graphémique *è*, *ê*, *ai*, *aî* ou *ei* quand il est en fin de syllabe et qu'il est précédé et suivi d'une consonne.

Exemples : *scène*, *règle*, *fête*, *capitaine*, *baigne*, *maître*, *reine*, *peigne*.

Cette règle compte, en moyenne, vingt-cinq occurrences par étudiant. Son taux de réussite, présenté en fonction des variables indépendantes *niveau d'études* et *rendement en français*, apparaît au tableau 3.24. Il ressort, de l'examen de ces résultats, que la règle R6 est déjà bien maîtrisée dès la deuxième année, le taux moyen de réussite étant de 78,0% pour ce niveau. On remarque aussi qu'entre la deuxième et la sixième année, il n'y a pas de progrès significatif, le taux de réussite passant de 78,0% à 80,6%. Par contre, sauf pour la quatrième année, à tous les autres niveaux, les forts ont un taux de réussite clairement plus élevé que celui des faibles. Les étudiants forts de deuxième année, avec un taux de réussite de 85,8%, possèdent une maîtrise de la règle R6 que les étudiants faibles n'atteindront pas à aucun des niveaux d'études du cours primaire. Le taux de réussite le plus élevé chez les étudiants faibles est celui des étudiants de cinquième

TABLEAU 3.24

TAUX MOYENS DE RÉUSSITE DE LA RÈGLE R6
EN FONCTION DU NIVEAU D'ÉTUDES ET
DU RENDEMENT EN FRANCAIS

NIVEAUX	RENDEMENTS		
	FORTS %	FAIBLES %	MOYENNE %
2	85,8	70,1	78,0
3	86,5	73,3	79,9
4	81,8	80,6	81,2
5	86,0	81,0	83,5
6	84,0	77,1	80,6
Moyenne	84,8	76,4	80,6

année qui ont obtenu une moyenne de 81,0%. Cette avance des forts sur les faibles se reflète dans le taux de réussite moyen qui est de 84,8% pour l'ensemble des forts et de 76,4% pour l'ensemble des faibles.

Une analyse de variance effectuée à partir des taux de réussite à la règle R6 et dont les résultats sont présentés au tableau 3.25 indique que la variable *rendement en français* a un effet sur le taux de réussite à cette règle. La variable *niveau d'études* n'a pas, quant à elle, d'effet significatif, la règle R6 étant déjà assez bien maîtrisée dès la deuxième année et les progrès, pratiquement nuls de la deuxième à la sixième année. L'écart entre le taux de réussite des forts et des faibles étant plutôt constant d'un niveau à l'autre, l'interaction *niveau d'études - rendement en français* n'est pas significatif.

TABLEAU 3.25

RÉSULTATS DE L'ANALYSE DE VARIANCE
EFFECTUÉE À PARTIR DES RÉSULTATS OBTENUS
À LA RÈGLE R6

SOURCES DE VARIATION	CARRÉS MOYENS	DEGRÉS DE LIBERTÉ	F	NIVEAU DE SIGN.
Niveau d'études	110,9	4	0,41	0,800
Rendement en français	3363,9	1	12,49	0,001
Interaction niv.-rend.	322,2	4	1,19	0,314
Résidus	269,4	183		

3.4.7 Conclusion

L'hypothèse E3 pose que l'individu qui apprend à orthographier le français acquiert progressivement la maîtrise des règles de formation des suites graphémiques qui répondent aux normes de cette langue. Cette acquisition se ferait, au moins dans certains cas, vraisemblablement inconsciemment puisque les méthodes actuelles d'enseignement de l'orthographe lexicale ne s'arrêtent pas explicitement sur plusieurs des règles de formation des suites graphémiques en français.

Les résultats que les sujets de l'échantillon ont obtenus pour chacune des six règles étudiées démontrent que l'acquisition des règles est un phénomène complexe qui varie d'ailleurs d'une règle à l'autre. En effet, il existe certaines règles qui sont bien maîtrisées dès la deuxième année et pour l'acquisition desquelles il n'y a pratiquement pas d'évolution pendant les autres années du cours primaire. C'est le cas, par exemple, des règles R3 et R4 et dans une moindre mesure, de la règle R6. D'autres règles, comme R1 et R2, demeurent toujours relativement mal maîtrisées puisque même en sixième année, le taux de réussite demeure inférieur à 60%. Pour d'autres règles, enfin, on observe un progrès relativement continu de la deuxième à la sixième année. C'est le cas, par exemple, de la règle R5.

Sur les résultats obtenus aux six règles étudiées, on peut encore faire les observations suivantes. D'abord, on remarque qu'à la fin du cours primaire, les règles ne sont jamais encore parfaitement connues des élèves puisqu'en sixième année, les résultats varient de 52% à 90%. Il y a donc certainement lieu de tenter d'améliorer l'enseignement de l'orthographe lexicale au niveau primaire et possiblement de le poursuivre au niveau secondaire.

Il ressort encore, de l'ensemble des résultats portant sur le taux de réussite aux différentes règles, que pour un même niveau d'études, le taux de réussite des étudiants faibles en français est toujours inférieur à celui des étudiants forts; cette différence est d'autant plus grande que la règle est mal assimilée par les forts. Mis à part les règles R3 et R4 qui sont bien maîtrisées par les forts et les faibles dès la deuxième année, pour les autres règles, on remarque, de plus, que les étudiants forts affichent une avance sur les faibles qui varie de deux à quatre ans. Pour les règles R1 et R2 en effet, les étudiants forts de troisième année obtiennent de meilleurs scores que les étudiants faibles de sixième année. Pour la règle R5, le score des étudiants forts de deuxième année est supérieur à celui des étudiants faibles de quatrième année. C'est pour la règle R6 que l'écart est le plus grand puisque le score obtenu par les étudiants forts de deuxième année est supérieur à celui obtenu par les étudiants faibles de sixième année. Si l'on

en juge par l'importance de ces écarts entre les forts et faibles, il semble que la maîtrise des règles de formation des suites graphémiques caractérise assez fortement les étudiants dont le rendement en français est supérieur.

On observe également qu'à l'exception de la règle R1, les progrès les plus importants dans la maîtrise des règles sont réalisés vraisemblablement au niveau de la première année puisque déjà, en deuxième année, le taux de réussite atteint de 42,1% à 93,4% selon la règle. Sauf pour la règle R5, un autre progrès important se produit entre la deuxième et la troisième année. Ce progrès est toutefois beaucoup plus marqué chez les étudiants forts qui atteignent en fait leurs meilleures performances en troisième année. Il semble donc exister un plafonnement au taux de réussite à ce niveau pour les étudiants forts. On peut se demander si ce plafonnement pourrait être attribuable à l'approche pédagogique utilisée pour enseigner l'orthographe lexicale ou si, pour certaines règles complexes, il n'existerait pas des éléments que les étudiants ne parviendraient pas à maîtriser au niveau du cours primaire.

Quant aux progrès des étudiants faibles, ils sont plus continus que ceux des forts. A l'exception des règles R3 et R4 qui sont maîtrisées dès la deuxième année, on observe en effet un taux de réussite croissant de la deuxième à la cinquième année pour la règle R6 et de la deuxième à

la sixième année pour les autres règles.

L'ensemble des résultats obtenus pour les règles qui ne sont pas maitrîsées dès la deuxième année semble donc confirmer l'hypothèse E3 puisque jusqu'à la troisième année chez les forts, et jusqu'à la cinquième ou sixième année chez les faibles, on observe un taux de réussite croissant pour les règles de formation des suites graphémiques correctes en français. Cependant, comme il a déjà été mentionné lors de l'étude des taux de réussite aux règles R3 et R4, il est concevable que les sujets obtiennent de meilleurs scores pour la réussite des règles qui soient attribuables, non pas à une connaissance de ces règles mais à l'utilisation d'une autre stratégie qui expliquerait ces résultats. En effet, il serait possible que les individus utilisent les graphèmes en reproduisant inconsciemment les fréquences des diverses expressions graphémiques qui , en français, peuvent représenter un phonème. En utilisant les graphies dans des proportions qui se rapprochent de leur fréquence théorique en français, ils augmentent, dans une épreuve orthographique du type de celle qui leur a été présentée, leurs chances de succès aux règles.

Or la vérification de l'hypothèse E1 a permis de découvrir que le code des sujets de l'échantillon se rapproche progressivement du code français. Cette constatation signi-

fie que les fréquences empiriques des diverses graphies se rapprochent, avec les niveaux d'études, de leur fréquence théorique en français. Similairement, la maîtrise progressive des règles implique, jusqu'à un certain point, une convergence des fréquences d'utilisation des expressions graphémiques avec les fréquences dans lesquelles elles apparaissent dans la langue française.

Pour déterminer si c'est le rapprochement des fréquences utilisées et des fréquences théoriques qui explique la réussite lors de l'application des règles ou si c'est l'apprentissage des règles qui explique que la fréquence d'utilisation des graphies converge vers les fréquences théoriques en français, on a procédé au raisonnement suivant. Si les sujets de l'échantillon ne connaissent pas une règle donnée, la proportion des diverses graphies qu'ils ont utilisées dans les contextes décrits par cette règle pour représenter un phonème doit tendre vers la proportion des diverses graphies qu'ils ont utilisées dans l'épreuve orthographique pour représenter les occurrences de ce phonème.

Par exemple, si les contextes nécessitant l'utilisation des graphies ç ou ss sont au nombre de dix et si un sujet de l'échantillon utilise, dans 20% des représentations du phonème /s/ de l'épreuve orthographique, les graphies en question, alors, statistiquement, 20% (ou 2 sur 10) des contextes concernés devraient recevoir les graphies ç ou ss.

Si on retrouve, dans ces dix contextes, une proportion de ces graphies significativement plus élevée que 20%, on devra conclure que quelque chose d'autre que le hasard a amené l'étudiant à utiliser les graphies ç et ss dans ces contextes. Par contre, si la proportion des graphies ç et ss utilisées par l'étudiant dans ces dix environnements n'est pas sensiblement différente de 20%, on ne peut savoir, dans ce cas, s'il y a intégration des fréquences ou maîtrise de la règle.

Le tableau 3.26 présente, pour chacune des règles retenues pour la vérification de l'hypothèse E3, d'une part les proportions, exprimées en pourcentages, dans lesquelles les sujets de l'échantillon ont utilisé les graphies exigées pour la réussite de chaque règle. Cette information apparaît dans la colonne "Fréquences empiriques". D'autre part, apparaissent les taux de réussite des sujets de l'échantillon aux diverses règles. Cette information est contenue dans la colonne "Taux de réussite". Suivant le raisonnement qui vient d'être exposé, quand le taux de réussite à une règle donnée est sensiblement plus élevé que celui qui représente la fréquence d'utilisation dans la dictée des diverses graphies exigées pour la réussite de cette règle, on peut conclure que le taux de réussite reflète un degré de maîtrise de la règle.

Il ressort, de l'examen des données présentées au ta-

TABLEAU 3.26

COMPARAISON DES FRÉQUENCES EMPIRIQUES DES GRAPHIES EXIGÉES POUR LA RÉUSSITE DES RÈGLES ET DES TAUX DE RÉUSSITE OBTENUS À CES RÈGLES

NIVEAUX	RENDEMENTS	R1		R2		R3		R4		R5		R6	
		F.E.*	T.R.**	F.E.	T.R.	F.E.	T.R.	F.E.	T.R.	F.E.	T.R.	F.E.	T.R.
2	Forts	6,7	18,6	32,3	41,9	89,8	95,3	66,0	91,7	57,4	75,9	48,9	85,8
	Faibles	5,2	16,5	33,1	42,3	89,9	91,5	64,5	85,9	58,5	66,1	52,6	70,1
3	Forts	27,1	63,6	53,0	67,2	70,4	94,3	45,8	92,8	68,2	76,3	48,7	86,5
	Faibles	13,0	39,8	43,8	50,6	83,4	93,9	54,4	91,6	70,7	68,8	46,3	73,3
4	Forts	25,7	64,8	53,8	63,4	71,8	93,3	45,5	85,6	67,8	73,7	48,1	81,8
	Faibles	15,0	30,6	43,4	45,4	80,7	91,2	54,5	88,4	69,8	64,8	56,8	80,6
5	Forts	26,2	62,4	52,6	57,7	71,3	97,9	46,2	95,4	70,1	81,4	47,5	86,0
	Faibles	21,7	35,4	43,5	48,7	73,6	91,9	54,1	84,6	74,8	81,0	48,4	81,0
6	Forts	31,0	57,5	51,5	52,0	67,1	92,4	47,6	93,1	71,0	86,7	44,9	84.0
	Faibles	30,1	56,7	48,4	57,6	47,2	87,5	50,3	86,3	73,8	82,1	45,0	77,1

* F.E.: fréquence empirique

** T.R.: taux de réussite

bleau 3.26, que dans le cas des règles R1, R4 et R6, les fréquences empiriques sont nettement inférieures aux taux de réussite. Cette observation permet de conclure que les taux de réussite reflètent un degré de maîtrise de ces règles. Dans le cas de la règle R3, l'écart entre les fréquences empiriques et les taux de réussite n'est pas aussi grand, particulièrement pour les étudiants forts et faibles de deuxième année. On ne peut donc pas connaître, pour ces étudiants , le degré réel de maîtrise de cette règle. Pour les autres, par contre, l'écart entre les fréquences empiriques et les taux de réussite est peut-être suffisamment grand pour que l'on puisse penser qu'il y a maîtrise de la règle R3. Enfin, dans le cas des règles R2 et R5, l'écart entre les fréquences empiriques et les taux de réussite est petit pour trop de groupes d'étudiants pour qu'il soit possible de se prononcer avec un degré de certitude suffisant sur la maîtrise de ces règles.

Ce test qui vient d'être effectué de même que l'examen de la progression des taux de réussite aux diverses règles étudiées pour la vérification de l'hypothèse E3 permettent de conclure qu'il y a effectivement maîtrise progressive de certaines règles de formation des suites graphémiques correctes en français. Cette conclusion permet d'expliquer, au moins partiellement, le rapprochement entre le code des sujets de l'échantillon et le code français tel qu'il a été observé lors de la vérification de l'hypothèse E1. On peut

illustrer cette causalité dans le rapprochement des codes à l'aide, par exemple, des règles R1 et R2.

L'examen des taux de réussite obtenus aux règles R1 et R2 permet en effet de mieux comprendre le saut observé dans le rapprochement entre le code des étudiants et le code français pour le phonème /s/ entre la deuxième et la troisième année.

Les valeurs prises par l'indice d'utilisation des graphies obtenues lors de la vérification de l'hypothèse E2 indiquaient qu'en deuxième année, les graphies ss et s sont utilisées dans des proportions qui s'éloignent largement de leur fréquence théorique correspondante. Par ailleurs, l'examen des taux de réussite aux règles R1 et R2 montre que les taux subissent un progrès marqué entre la deuxième et la troisième année. Or, la graphie ss est l'une des graphies exigées pour la réussite de chacune de ces règles tandis que la graphie s est l'une de celles qui sont considérée comme mauvaises. Un progrès dans le taux de réussite de ces règles a donc normalement entraîné un accroissement de l'utilisation de la graphie ss qui était sous-utilisée en deuxième année et une diminution de l'utilisation de la graphie s qui était sur-utilisée à ce niveau. De telles variations dans l'utilisation de ces graphies expliquent comment la fréquence empirique de ss qui était d'environ 6% en deuxième année est passée à environ 14% en

troisième année, ce qui la rapproche de sa fréquence théorique qui est de 32,0%. De la même façon, ces variations expliquent comment la fréquence empirique de s qui était d'environ 65% en deuxième année est passée à environ 40% en troisième année, ce qui la rapproche, elle aussi, de sa fréquence théorique qui est de 32,4%.

Puisque les fréquences fournies par l'indice d'utilisation des graphies sont une composante de l'indice de similitude, un rapprochement des fréquences empirique et théorique explique la convergence qui a été observée entre le code des sujets de l'échantillon et le code français.

La confirmation, pour certaines règles, de l'hypothèse E3 est certainement intéressante puisqu'elle fournit des indications sur la façon dont se fait, au moins en partie, l'acquisition de l'orthographe lexicale. Pour obtenir de meilleures informations sur cette hypothèse, il aurait toutefois été préférable d'inclure, dans l'échantillon, des sujets plus jeunes que l'on aurait soumis à l'épreuve orthographique plus d'une fois au cours de leur première année de scolarité puisque les progrès les plus importants sont observés avant la deuxième année. De plus, à cause des nombreux types d'évolution des taux de réussite pour les différentes règles, il aurait également été préférable de vérifier cette hypothèse au moyen de règles plus fines et plus nombreuses.

Si l'évolution des valeurs prises par l'indice de similitude n'est pas attribuable à une maîtrise progressive des correspondances phonème-graphème, on a démontré qu'elle s'explique par contre, au moins partiellement, par une maîtrise progressive des règles de formation des suites graphémiques en français.

On a vu en effet que l'ensemble des résultats obtenus pour les règles qui ne sont pas maîtrisées dès la deuxième année confirme l'hypothèse E3 puisque jusqu'à la troisième année chez les forts, et jusqu'à la cinquième ou sixième année chez les faibles, on observe un taux de réussite croissant pour les règles de formation des suites graphémiques correctes en français.

Dans la section qui suit, on examinera s'il y a, chez les étudiants, une intégration des fréquences théoriques des diverses expressions graphémiques pouvant représenter un phonème donné et on tentera d'identifier son rôle dans l'acquisition du code chez les individus qui apprennent à orthographier le français.

3.5 Vérification de l'hypothèse E4

HYPOTHÈSE E4: L'individu qui apprend à orthographier le français sur-utilise, pour représenter un phonème, les expressions graphémiques les plus fréquentes de ce phonème.

Celui qui maîtrise les règles de formation des suites graphémiques correctes en français utilise, de ce fait, les diverses expressions graphémiques qui peuvent représenter les phonèmes de cette langue dans des proportions qui correspondent à celles du français standard. Or, on a vu, à la section 3.4, que tandis que certaines règles de formation de suites graphémiques sont bien maîtrisées dès la deuxième année, d'autres ne le sont pas encore en sixième année. On peut donc se demander quelles expresssions graphémiques choisira, pour écrire des logatomes, un individu qui ne connaît pas certaines règles de formation des suites graphémiques. Pour représenter un phonème, l'individu fera-t-il appel à plusieurs expressions graphémiques correspondant à ce graphème ou se concentrera-t-il sur une des graphies qu'il connaît?

On a conjecturé que dans une telle situation, un individu aurait naturellement recours aux expressions graphémi-

ques les plus fréquentes en français, c'est-à-dire, celles qu'il aurait vues le plus souvent et donc, qu'il connaîtrait le mieux. Pour vérifier cette hypothèse, on a développé un instrument de mesure, le rapport fréquence empirique - fréquence théorique qui a été calculé à partir du dépouillement de l'épreuve orthographique des sujets de l'échantillon qui ont obtenu, aux règles énoncées à la section précédente, un taux de réussite inférieur.

3.5.1 Le rapport fréquence empirique - fréquence théorique

On connaît les fréquences, en français, de chacune des expressions graphémiques utilisées pour représenter les phonèmes /s/, /ɛ/ et /o/ en position médiane. Ces fréquences, présentées aux tableaux 2.01, 2.02 et 2.03, sont ce qu'on a convenu d'appeler les fréquences théoriques. Quant aux fréquences empiriques, ce sont celles qui ont été utilisées par les sujets de l'échantillon pour représenter, dans l'épreuve orthographique, les phonèmes étudiés. Le calcul du rapport fréquence empirique - fréquence théorique pour chacune des expressions graphémiques les plus utilisées en français permet de savoir si les graphies auxquelles les sujets de l'échantillon ont eu le plus souvent recours pour représenter les phonèmes /s/, /ɛ/ et /o/ sont celles qui sont les plus fréquentes dans la langue française. Lorsque, pour une expression graphémique donnée, le rapport est supérieur à un, on sait que cette graphie a été utilisée plus fréquemment dans la transcription des logatomes qu'elle n'aurait dû l'être si le code français avait été rigoureusement respecté. Un rapport inférieur à un indique une sous-représentation d'une expression graphémique.

L'objectif de cette section est d'étudier quelles sont les graphies utilisées par ceux qui ne connaissent pas les règles de formation des suites graphémiques françaises pour écrire des logatomes. Conséquemment, seuls les sujets de

l'échantillon qui ont eu un score faible aux règles des phonèmes /s/ et /ɛ/ traitées à la section 3.4 ont été retenus pour les fins de cette analyse. Pour le phonème /s/, les individus qui ont obtenu, aux règles R1 et R2, une moyenne inférieure à un pourcentage qu'on a arbitrairement fixé à 70% sont ceux qui ont été retenus. Ils représentent 65% de l'ensemble des sujets de l'échantillon. Les scores obtenus aux règles R3 et R4 n'ont pas été pris en considération puisque ces règles ont été réussies par à peu près tous les sujets de l'échantillon. Puisque dans l'ensemble, les règles relatives au phonème /ɛ/ ont été mieux réussies que celles du phonème /s/, ce sont, cette fois, les individus qui ont eu un score moyen inférieur à un pourcentage arbitrairement fixé à 75% pour les règles R5 et R6 qui ont été retenus. Ces individus représentent 45% de l'ensemble des sujets de l'échantillon.

3.5.1.1 Le rapport fréquence empirique - fréquence théorique pour le phonème /s/

Les rapports fréquence empirique - fréquence théorique pour le phonème /s/ sont présentés au tableau 3.27. On notera que ces rapports sont présentés uniquement en fonction de la variable *niveau d'études*. Les règles ayant généralement été mieux réussies par les étudiants dont le rendement en français est supérieur, une analyse en fonction de la variable *rendement en français* n'aurait pu être très valable puisque le groupe des forts aurait été sous-représenté.

L'expression graphémique à laquelle correspond le rapport fréquence empirique - fréquence théorique le plus élevé est *s*. Ce rapport a une valeur moyenne de 1.72. Or, cette graphie qui, en français représente le phonème /s/ dans une proportion de 32,4%, est la plus fréquente des graphies utilisées pour représenter ce phonème. La graphie *c* à laquelle correspond une fréquence théorique de 22,7% possède aussi un rapport fréquence empirique - fréquence théorique élevé. Sa valeur est de 1,07. Si on ne considérait que ces deux expressions graphémiques, l'hypothèse E4 serait confirmée. Avec des rapports fréquence empirique - fréquence théorique supérieurs à l'unité, ces deux graphies, auxquelles correspondent des fréquences théoriques élevées, sont utilisées dans des proportions encore plus élevées par les sujets de l'échantillon qui ne maîtrisent pas bien les règles de for-

TABLEAU 3.27

RAPPORTS FRÉQUENCE EMPIRIQUE - FRÉQUENCE THÉORIQUE POUR LE PHONÈME /s/ EN FONCTION DU NIVEAU D'ÉTUDES

NIVEAUX	POURCENTAGES D'ÉTUDIANTS RETENUS	EXPRESSIONS GRAPHÉMIQUES				
		s	ss	c	ç	autres
2	71	2,06	0,19	0,89	0,02	0,48
3	57	1,56	0,23	1,39	0,80	0,33
4	75	1,64	0,42	1,14	0,30	0,36
5	70	1,62	0,48	1,11	0,36	0,40
6	55	1,74	0,58	0,83	0,40	0.24
Moyenne	65	1,72	0,38	1,07	0,38	0,36
Fréq. théoriques		32,4%	32,0%	22,7%	6,5%	6,4%

mation des suites graphémiques du phonème /s/.

Toutefois, il ressort, des données présentées au tableau 3.27, que la graphie ss à laquelle correspond une fréquence théorique qui est à peu près équivalente à celle de l'expression graphémique s, soit 32,0%, possède un rapport fréquence empirique - fréquence théorique dont la valeur n'est que de 0,38, donc nettement inférieure à l'unité. Cette dernière observation infirme donc l'hypothèse E4 au moins pour ce qui est du phonème /s/. Les individus qui ne maîtrisent pas les règles de formation des suites graphémiques du phonème /s/ n'utilisent pas systématiquement plus fréquemment, pour représenter ce phonème, les expressions graphémiques les plus fréquentes en français.

On aurait pu croire que la faible valeur du rapport fréquence empirique - fréquence théorique de l'expression graphémique ss soit dû au fait que les écoliers, particulièrement les plus jeunes, n'aient pas perçu les graphies s et ss comme deux graphies différentes et aient indifféremment eu recours à l'une ou à l'autre. En sixième année, par contre, il serait étonnant que la distinction entre ces deux graphies n'ait pas encore été assimilée. Pourtant, le rapport fréquence empirique - fréquence théorique de la graphie ss, à ce niveau, n'est que de 0,58. Il semble donc que la graphie ss soit réellement sous-utilisée et que l'on ait raison

de conclure que l'hypothèse E4 est infirmée, au moins pour ce qui concerne le phonème /s/.

Lors de la vérification de l'hypothèse E3, on avait remarqué que les étudiants forts et faibles de sixième année auraient dû obtenir un score plus élevé que celui qu'ils ont obtenu pour les règles R1 et R2. En effet, les graphies exigées pour la réussite de ces règles, surtout dans le cas des plus fréquentes comme ss et c, ont été utilisées, par les sujets de l'échantillon de sixième année, dans des proportions qui se rapprochent sensiblement de leurs fréquences théoriques. Toutefois, si on ne considère que le groupe d'étudiants de ce niveau qui ont obtenu un score inférieur aux règles du phonème /s/, donc aux règles R1 et R2, leur rapport fréquence empirique - fréquence théorique révèle que certaines expressions graphémiques ont été largement sous-utilisées. La graphie ss, par exemple, qui est exigée pour la réussite de chacune des règles R1 et R2 a un rapport qui n'est que de 0,58. Parce que ce groupe représente 55% des sujets de l'échantillon de sixième année, on comprend mieux comment il se fait que l'ensemble des étudiants forts et faibles de ce niveau n'avaient obtenu qu'un taux de réussite situé aux environs de 50% pour ces deux règles.

3.5.1.2 Le rapport fréquence empirique - fréquence théorique pour le phonème /ɛ/

Les rapports fréquence empirique - fréquence théorique pour le phonème /ɛ/ sont présentés au tableau 3.28. Comme pour le phonème /s/, ces rapports sont présentés uniquement en fonction de la variable *niveau d'études*.

Avec une fréquence théorique de 51,3%, la graphie *e* est, parmi les expressions graphémiques qui peuvent représenter le phonème /ɛ/ en position médiane, celle qui a la fréquence théorique la plus élevée. Puisque son rapport moyen fréquence empirique - fréquence théorique, qui est de 0,96%, se rapproche de l'unité, on en déduit que la fréquence d'utilisation de cette graphie, par les sujets de l'échantillon qui maîtrisent mal les règles du phonème /ɛ/, se rapproche sensiblement de sa fréquence théorique. L'expression graphémique *ai* est également souvent utilisée, en français, pour représenter le phonème /ɛ/. Sa fréquence théorique est en effet de 20,8%. Elle est cependant très peu utilisée par les écoliers de deuxième année qui maîtrisent mal les règles du phonème /ɛ/. Ceux des autres niveaux l'utilisent dans un rapport qui se rapproche de l'unité.

Les deux expressions graphémiques *e* et *ai* ne sont donc pas utilisées avec des fréquences plus élevées que leurs fréquences théoriques comme le voudrait l'hypothèse E4. En

TABLEAU 3.28

RAPPORTS FRÉQUENCE EMPIRIQUE - FRÉQUENCE THÉORIQUE POUR LE PHONÈME /ɛ/ EN FONCTION DU NIVEAU D'ÉTUDES

NIVEAUX	POURCENTAGES D'ÉTUDIANTS RETENUS	EXPRESSIONS GRAPHÉMIQUES				
		e	*ai*	*è*	*ê*	*autres*
2	46	0,84	0,27	2,83	0,25	0,35
3	55	1,00	0,97	1,14	0,49	0,48
4	62	0,91	1,06	1,65	0,31	0,57
5	35	1,07	0,73	1,57	0,28	0,50
6	28	0,99	1,01	1,56	0,42	0,42
Moyenne	45	0,96	0,81	1,75	0,35	0,46
Fréq. théoriques		51,3%	20,8%	13,1%	6.5%	8.3%

fait, leurs fréquences théoriques sont sensiblement bien respectées. Par contre, avec une fréquence théorique qui n'est que de 13,1%, la graphie *ê* possède unrapport fréquence empirique - fréquence théorique de 1,75. Ainsi, cette graphie qui ne représente pas très souvent le phonème /ɛ/ en français est utilisée plus fréquemment qu'elle ne devrait. On déduit donc, de l'ensemble de ces données, que lorsque les étudiants ne savent pas laquelle des graphies du phonème /ɛ/ doit être utilisée, ils ne font pas systématiquement appel aux expressions graphémiques les plus fréquentes de ce phonème.

3.5.2 Conclusion

Comme l'indiquent les rapports des fréquences empiriques sur les fréquences théoriques pour les phonèmes /s/ et /ɛ/, ce ne sont pas toujours, pour un phonème donné, les graphies qui sont les plus fréquentes dans la langue française qui sont utilisées le plus souvent par les individus qui ne savent pas à quelle expression graphémique ils doivent avoir recours pour représenter ce phonème. Pourtant, l'utilisation des graphies les plus fréquentes maximiserait, pour un individu en situation d'incertitude, ses chances de succès.

Le fait que les individus ne recourent pas de façon systématique aux graphies les plus fréquentes incite à penser qu'ils ne sont pas conscients des fréquences théoriques d'utilisation des expressions graphémiques pour représenter les divers phonèmes. Une vérification informelle de cette hypothèse auprès d'adultes francophones indique que ceux-ci ne sont effectivment pas conscients du nombre ou de la fréquence des graphies que la langue française utilise pour représenter un phonème. Par exemple, sur vingt adultes interrogés, aucun n'a donné, comme graphie la plus fréquente pour représenter le phonème /ɛ/ en position médiane le e qui constitue pourtant plus de 50% des graphies de ce phonème. Dans ce cas, il n'est donc pas étonnant que les individus

qui ne savent pas comment représenter un phonème dans un mot donné ne se réfèrent pas aux expressions graphémiques les plus fréquentes de ce phonème.

L'infirmation de l'hypothèse E4, qui découle du fait que les rapports fréquence empirique-fréquence théorique pour les graphies les plus fréquentes des phonèmes étudiés sont souvent inférieurs à l'unité, n'implique cependant pas que pour représenter un phonème donné, les étudiants en situation d'incertitude choisissent des graphies au hasard. Il est en effet possible qu'ils aient, consciemment ou non, recours à d'autres stratégies que la présente étude n'explore pas.

CONCLUSION

Il existe au moins deux raisons pour lesquelles la maîtrise de l'orthographe est importante. D'abord la conformité des écrits d'un individu aux normes établies en matière d'orthographe facilite la communication en favorisant la vitesse de lecture et en réduisant les ambiguïtés. De plus, la maîtrise de l'orthographe est utilisée, par la société, comme critère pour catégoriser les individus. Une personne dont l'orthographe est incorrecte devient, pour ceux qui lisent ses écrits et qui connaissent le code de cette langue, une personne dont on peut douter des habiletés intellectuelles.

A cause de cette importance que représente la maîtrise de l'orthographe, l'école a centré une grande partie de ses énergies sur l'acquisition de la norme orthographique et de la grammaire de la langue écrite. Elle n'a t utefois pas, pour autant, solutionné le problème puisque malgré le recours à des méthodes d'apprentissage variées et qui semblent de mieux en mieux adaptées à l'enfant, l'orthographe fran-

çaise n'est toujours pas maîtrisée par une grande partie des individus au moment où ils se présentent sur le marché du travail.

Le problème de l'apprentissage de l'orthographe française se pose depuis toujours et notamment depuis que celle-ci a été fixée par l'Académie française en 1835. Il se peut qu'on ne puisse jamais y apporter de solution, simplement parce que le code français serait trop complexe pour être parfaitement maîtrisé par la plupart des francophones. La difficulté que présente l'apprentissage de l'orthographe peut aussi provenir du fait que l'on ne possède pratiquement aucune information sur les mécanismes qui interviennent réellement dans cet apprentissage.

La présente étude a voulu contribuer à pallier ce manque d'information en s'intéressant particulièrement à l'acquisition de l'orthographe lexicale. Essentiellement, on a tenté de comprendre comment un individu acquiert la maîtrise de l'orthographe lexicale.

Pour atteindre l'objectif fixé, on s'est d'abord arrêté sur quelques caractéristiques de l'orthographe française ainsi que sur la perception qu'en ont quelques chercheurs qui s'y sont intéressés. On a vu que si quelques-uns n'y trouvent que désordre et anarchie et proposent, conséquemment, de la réformer, d'autres soutiennent que ce désordre n'est

qu'apparent et qu'une étude approfondie du phénomène permet de dégager des structures rationnelles dans l'orthographe française.

L'histoire montre que les arguments des seconds ont sûrement été mieux entendus que ceux des premiers puisque l'Académie Française a rejeté toutes les propositions de réforme de l'orthographe. Il ne reste donc plus, au pédagogue, que de prendre son parti de la situation et de développer des méthodes d'enseignement susceptibles de faciliter l'assimilation du code français. Aussi, les principales études qui ont porté sur l'apprentissage de l'orthographe lexicale ont-elles été passées en revue. Comme on l'a vu, la plupart de ces études ne font que fournir des vocabulaires de base et indiquer les endroits, à l'intérieur des mots, où les erreurs sont le plus souvent commises. Elles ne renseignent donc pas sur la façon dont se fait l'apprentissage de l'orthographe lexicale. Elles semblent supposer que celle-ci est acquise par la mémorisation de chacun des mots de la langue et cherchent des moyens pour favoriser une mémorisation efficace en tenant compte de variables telles l'intérêt ou la difficulté de chacun des mots.

Etant donné que, d'une part, il existe de nombreuses relations entre la lecture et l'écriture et que, d'autre part, les études qui ont porté sur l'apprentissage de la lecture ont poussé beaucoup plus loin l'analyse des mécanis-

mes d'apprentissage de cette habileté, il a semblé approprié de tenir compte des résultats de ces dernières pour mieux comprendre comment se fait l'acquisition de l'orthographe lexicale.

Les études sur la lecture indiquent que l'exercice de cette habileté impliquerait le recours à au moins deux types de connaissances. D'une part, le lecteur doit savoir associer une suite phonologique donnée à la suite graphémique correspondante. D'autre part, il doit pouvoir identifier les suites graphémiques correctes dans sa langue.

A la suite des études sur l'apprentissage de la lecture, il ne paraît pas invraisemblable de penser que celui qui apprend à écrire dans une langue suivrait un processus semblable à celui qui apprend à lire. On a ainsi été amené à énoncer quatre hypothèses de travail portant sur la façon dont se fait l'apprentissage de l'orthographe lexicale.

Pour vérifier la justesse de ces hypothèses, une épreuve orthographique constituée de logatomes a été construite et soumise à des étudiants de la deuxième à la sixième année du niveau primaire. La transcription de ces logatomes par chacun des sujets de l'échantillon a été codifiée et placée sur support informatique. L'analyse des suites graphémiques auxquelles les étudiants ont eu recours pour représenter les mots fictifs qui leur ont été dictés a été faite en tenant compte des variables indépendantes retenues pour les fins de

cette étude, soit le niveau d'études auquel un étudiant est inscrit et son rendement en français.

La première des quatre hypothèses qui ont été formulées dans le cadre du présent travail (hypothèse E1) pose que la maîtrise de l'orthographe lexicale ne se fait pas uniquement par la mémorisation de la représentation graphémique de chacun des mots de la langue. Pour vérifier cette hypothèse, on a défini une variable dépendante, l'indice de similitude au code français. Basée sur les fréquences des expressions graphémiques théoriquement utilisées en français pour représenter les phonèmes retenus pour les fins de cette étude, soit /s/, /ɛ/ et /o/, l'indice de similitude au code français sert à mesurer le degré de rapprochement du code d'un individu au code français.

L'examen des résultats démontre que les proportions dans lesquelles les étudiants utilisent les diverses expressions graphémiques pour représenter un phonème se rapprochent progressivement de celles du français à mesure que les étudiants accèdent à des niveaux de scolarité plus avancés. De plus, le degré de rapprochement du code des individus au code français est significativement plus élevé chez les étudiants dont le rendement en français est supérieur que chez ceux dont le rendement est plus faible. Ces résultats indiquent qu'il y aurait donc un apprentissage du code écrit qui ne serait clairement pas attribuable à la

mémorisation lexicale. En effet, si les individus apprenaient uniquement par la mémorisation de la représentation graphémique de chaque mot de la langue, il n'y aurait pas convergence entre leur code et celui du code français dans une tâche de reproduction de logatomes. L'hypothèse E1 est donc confirmée.

Pour expliquer cette convergence, trois possibilités ont été explorées et ont, chacune, fait l'objet d'une hypothèse de travail. La première (E2) pose que l'acquisition de l'habileté à orthographier peut s'expliquer par une maîtrise progressive des diverses représentations graphémiques des phonèmes d'une langue. La seconde (E3) affirme que la convergence peut être attribuée à une maîtrise progressive des règles de formation des suites graphémiques. Enfin, la troisième (E4) propose qu'il y a une intégration, consciente ou non, des fréquences des diverses expressions graphémiques en français.

Selon l'hypothèse E2, une personne qui apprend à orthographier le français acquiert progressivement la maîtrise des représentations graphémiques des phonèmes de cette langue. Pour vérifier cette hypothèse, on a construit trois outils de mesure complémentaires qui donnent des indications sur la connaissance que les sujets de l'échantillon ont des expressions graphémiques qui peuvent être associées à un phonème. Les résultats indiquent que dès la deuxième année

pour les phonèmes /s/ et /ɛ/ et dès la troisième année pour le phonème /o/, tous les sujets de l'échantillon, forts et faibles, savent associer à un phonème, les diverses expressions graphémiques qui peuvent le représenter en position médiane. La convergence du code des sujets de l'échantillon vers le code français (Hypothèse E1) ne peut donc être expliquée par une maîtrise progressive des correspondances phonèmes-graphèmes puisque les données recueillies lors de la vérification de l'hypothèse E2 indiquent qu'elles sont généralement acquises dès la deuxième année.

Cependant, un examen des valeurs prises par l'indice d'utilisation des graphies indique qu'en général, pour les graphies les plus fréquemment utilisées en français, on observe un rapprochement important des fréquences empirique et théorique entre la deuxième et la troisième année surtout chez les forts. Ce rapprochement est relativement plus lent au-delà de la troisième année. Quant aux graphies plus rares, leur convergence ne suit pas un modèle unique. L'indice d'utilisation des graphies permet d'analyser l'évolution de l'indice de similitude mais ne fournit pas d'indication sur les causes de cette évolution. La vérification de l'hypothèse E3, quant à elle, contribue à expliquer cette évolution.

Cette hypothèse pose qu'un individu qui apprend à écrire le français acquiert progressivement la maîtrise des

règles de formation des suites graphémiques qui répondent aux normes de cette langue. Pour la vérifier, des règles de formation de suites graphémiques ont été formulées. Les résultats que les sujets de l'échantillon ont obtenus pour chacune de ces règles démontrent que leur acquisition est un phénomène complexe qui varie de l'une à l'autre. En effet, il existe certaines règles qui sont bien maîtrisées dès la deuxième année et pour l'acquisition desquelles il n'y a pratiquement pas d'évolution pendant les autres années du cours primaire. Par ailleurs, d'autres règles demeurent toujours relativement mal maîtrisées puisque même en sixième année, le taux de réussite demeure inférieur à 60%. Enfin, pour certaines règles, on observe un progrès continu de la deuxième à la sixième année.

L'ensemble des résultats obtenus confirme tout de même l'hypothèse E3. En effet, quand une règle n'est pas déjà maîtrisée dès la deuxième année, les résultats indiquent que jusqu'en troisième année chez les forts et jusqu'en cinquième ou sixième année chez les faibles, il y a maîtrise progressive des règles de formation des suites graphémiques correctes en français.

En plus de confirmer l'hypothèse E3, ces résultats indiquent qu'il semble exister, pour les étudiants forts, un plafonnement au taux de réussite en troisième année et ce, même quand la règle n'est pas parfaitement maîtrisée.

On peut se demander si le plafonnement est attribuable à la méthode d'enseignement de l'orthographe lexicale. On peut aussi se demander ce qu'apprennent les étudiants forts au-delà de la troisième année. Consacrent-ils davantage d'énergie à l'acquisition de l'orthographe grammaticale? Cette étude ne peut apporter de réponse à ce genre de question. Elle permet tout de même d'observer que malgré un plafonnement en troisième année, les étudiants forts affichent une avance sur les étudiants faibles qui varie de deux à quatre ans selon la règle.

Evidemment, les étudiants qui maîtrisent les règles de formation des suites graphémiques correctes en français utilisent les diverses expressions graphémiques dans des proportions qui correspondent à celles du français standard. On peut cependant se demander à quelles graphies les étudiants qui ne maîtrisent pas ces règles auront recours pour représenter les phonèmes étudiés. L'hypothèse E4 pose que ces étudiants utilisent les expressions graphémiques les plus fréquentes en français.

Le rapport fréquence empirique - fréquence théorique utilisé pour vérifier cette hypothèse indique que ce ne sont pas toujours, pour un phonème donné, les graphies qui sont les plus fréquentes en français qui sont utilisées le plus souvent par les individus qui ne savent pas à quelle expression graphémique ils doivent avoir recours pour re-

présenter ce phonème. Le fait que l'hypothèse E4 soit infirmée incite à penser que les individus ne sont pas conscients des fréquences théoriques d'utilisation des expressions graphémiques pour représenter les divers phonèmes.

Au sujet de l'orthographe lexicale, Grevisse écrivait:

> L'orthographe lexicale... "s'acquiert, dans une certaine mesure, par la connaissance de l'étymologie, mais elle relève avant tout de l'esprit d'observation: mémoire visuelle, mémoire auditive, mémoire musculaire concourent à graver dans l'esprit *l'image de chaque vocable.*(1) Tel restera incurablement ignorant de l'orthographe aussi longtemps qu'il n'aura pas appris à observer la figure des mots"
> (Grevisse, 1975, p. 124)

Cette conception de la façon dont se fait l'apprentissage de l'orthographe lexicale était également celle de Buyse, Aristizabal (1938), Dubois (1939), Lambert (1947), Pirenne (1948), Ters (1977) et tous ceux qui ont, pour favoriser l'apprentissage de l'orthographe lexicale, établi des vocabulaires de mots présentés en fonction de l'intérêt qu'ils suscitent et du niveau de difficulté qu'ils présentent.

Or la vérification de l'hypothèse E1 a démontré qu'il y a un apprentissage du code écrit qui n'est clairement pas

---

(1) La mise en italiques est attribuable à l'auteur de cette étude.

attribuable à la mémorisation de la configuration graphémique des mots. Cette conclusion remet donc en question la conception que l'on a depuis longtemps de la façon dont se fait l'apprentissage de l'orthographe lexicale. Dans la mesure où les méthodes d'enseignement se fondent sur cette conception, il y aurait lieu de les ré-évaluer. Il est possible qu'on en arrive alors à élaborer de nouvelles approches qui tiendraient compte des conclusions tirées de la vérification des hypothèses E2 et E3. En effet, la vérification de l'hypothèse E2 a permis de se rendre compte que la convergence du code des sujets de l'échantillon vers le code français, qui évolue de la deuxième à la sixième année, ne peut être attribuée à une maîtrise progressive des associations phonème-graphème puisque, comme on l'a vu, ces associations sont généralement maîtrisées dès la deuxième année. Il serait donc approprié de ne pas insister sur l'acquisition de cette habileté au-delà de la deuxième année. Un retard dans l'acquisition de l'orthographe lexicale ne pourrait semble-t-il être attribué qu'exceptionnellement à une incapacité à associer les diverses expressions graphémiques aux phonèmes correspondants.

Quant aux résultats obtenus lors de la vérification de l'hypothèse E3, ils indiquent que c'est au moins en partie par l'acquisition des règles de formation des suites graphémiques correctes en français que se ferait l'acquisition de la maîtrise de l'orthographe lexicale. A la suite de cette constatation, il semble logi-

que de préconiser une méthode d'enseignement de l'orthographe lexicale qui tiendrait compte du fait que les enfants assimilent de telles règles. Cependant, il importe de se rappeler que ces règles ne sont pas déterministes dans le sens où, à un phonème donné, ne pourrait correspondre qu'une seule réalisation graphémique dans un environnement phonographique déterminé. On peut penser, par exemple, aux mots homophones *mer*, *mère*, *maire*, ou *vers*, *vert*, *verre*. Une méthode d'apprentissage de l'orthographe française devra donc comporter aussi une composante de mémorisation.

Il existe plusieurs façons possibles d'enseigner les règles de formation des suites graphémiques correctes en français. Une façon est de les enseigner de manière explicite. Une autre est d'exposer l'enfant à des séquences de mots qui ont en commun un phonème dont l'environnement varie d'un mot à l'autre tout en étant soumis à un nombre restreint de règles.

La première méthode pourrait être appropriée pour les étudiants qui ont déjà commencé leur apprentissage de l'orthographe depuis quelques années. Elle leur permettrait d'acquérir une connaissance plus systématique du code écrit de leur langue. Les résultats obtenus par la vérification des quatre hypothèses de travail de la présente étude portent toutefois à penser que pour les enfants en début d'apprentissage, la seconde méthode serait nettement plus appropriée.

En effet, puisqu'on a pu démontrer que des règles qui ne sont probablement pas enseignées sont apprises par les enfants, on en déduit que cet apprentissage est dû à de nombreuses expositions à ces règles. Or, dans les textes qui leur sont destinés, les mots sont généralement sélectionnés d'abord en fonction de l'intérêt qu'ils peuvent susciter et du niveau de difficulté qu'ils présentent. La fréquence d'exposition aux diverses règles de formation des suites graphémiques est donc laissée au hasard. Pourtant, les enfants réussissent, malgré tout, à les assimiler au moins en partie. A plus forte raison les apprendraient-ils plus facilement et plus rapidement si les textes auxquels ils sont exposés présentaient ces règles de façon méthodique.

Il n'est pas certain que cette stratégie particulière permettrait d'atteindre des résultats optimaux. Ce qui est plus sûr, par contre, c'est que l'approche pédagogique utilisée pour promouvoir l'acquisition de l'orthographe lexicale devra tenir compte du fait que l'enfant apprend, sans qu'on les lui enseigne nécessairement, les règles de formation des suites graphémiques correctes en français. Des recherches ultérieures devraient tenter d'évaluer diverses stratégies d'enseignement de l'orthographe qui répondent à cette réalité.

La présente étude a établi que l'acquisition des

règles de formation des suites graphémiques correctes en français est un phénomène complexe. En effet, tandis que certaines règles sont assez bien maîtrisées dès la deuxième année, d'autres ne le sont encore que très mal à la fin du cours primaire. De plus, pour le cas de certaines règles, il semble exister un plafonnement que ni les étudiants forts ni les étudiants faibles ne semblent pouvoir dépasser au moins au cours des six premières années de scolarité. La présente recherche n'explique pas ce phénomène. Pour le comprendre et pour mettre au point une méthode d'enseignement appropriée, il serait nécessaire qu'une étude plus élaborée soit entreprise. Cette étude devrait porter sur un plus grand nombre de phonèmes afin de vérifier si les conclusions qui vaïent pour les trois phonèmes étudiés peuvent être généralisées aux autres phonèmes. Cette étude devrait aussi contrôler des règles plus fines et plus nombreuses. Il serait peut-être possible, de cette façon, de comprendre ce qui caractérise les règles qui suivent un même type d'acquisition.

La présente étude a démontré que la maîtrise des associations phonème-graphème est généralement acquise très tôt chez les enfants, vraisemblablement au cours de la première année d'études. De même, la maîtrise des règles de formation des suites graphémiques correctes en français commence très tôt. Le progrès le plus marqué semble en effet

se produire avant la deuxième année. A cause de la précocité dans le développement de ces habiletés, il serait sans doute préférable d'inclure, dans un futur échantillon, des enfants plus jeunes.

De plus, puisque pour certaines règles, les progrès sont importants d'une année à l'autre, il serait souhaitable que les sujets de l'échantillon soient soumis à l'épreuve orthographique plus d'une fois au cours d'une année scolaire. Cette modalité permettrait en effet de suivre de plus près les progrès dans l'apprentissage des règles.

Pour mieux comprendre l'évolution de l'habileté à orthographier, il serait également préférable qu'une étude ultérieure tienne davantage compte de l'enseignement reçu par les étudiants.

La présente recherche a été effectuée auprès d'un échantillon provenant d'un milieu socio-économique plutôt aisé. Dans la mesure où les habiletés à lire et à écrire se développent plus lentement chez les enfants des groupes sociaux plus défavorisés, il est probable qu'une étude semblable menée auprès d'enfants de ces milieux conduise à des résultats quelque peu différents. Les conclusions de cette étude ne peuvent donc pas être généralisées sans ré-

serve à toute la population des classes du primaire.

En dépit de ses nombreuses limites dues entre autres au nombre restreint de phonèmes et de règles étudiés et au choix de l'échantillon, cette étude ouvre malgré tout des voies d'exploration dans un domaine peu connu. La principale ligne de force qui s'en dégage est que les enfants apprennent à maîtriser le code phono-graphique français même quand ce code ne semble pas faire l'objet d'un enseignement explicite. Il reste encore à découvrir les facteurs susceptibles de favoriser cet apprentissage. Une fois ces facteurs identifiés, il sera possible de concevoir et d'expérimenter des méthodes qui favoriseront la maîtrise du code phono-graphique, donc, en partie, de l'orthographe lexicale.

GLOSSAIRE

CODE PHONO-GRAPHIQUE : *L'ensemble des règles de correspondance entre graphèmes et phonèmes pratiquées par un individu, un milieu ou une époque.*

EXPRESSION GRAPHÉMIQUE : *Une seule lettre ou plusieurs lettres qui se suivent et qui représentent un seul phonème.*

GRAPHÈME : *Elément qui correspond, dans l'écriture alphabétique, à la lettre.*

LOGATOME : *Une syllabe ou une suite de syllabes appartenant à une langue mais ne formant pas un mot ou un syntagme significatif dans cette langue.*

PHONÈME : *Dans la chaîne parlée, unité distinctive minimale. ex: /p/vs/b/*

SUITE GRAPHÉMIQUE : *Ensemble de graphèmes qui se suivent et qui représentent plus d'un phonème.*

RÈGLE : *Formule qui indique quelles expressions graphémiques peuvent être utilisées dans un contexte donné.*

BIBLIOGRAPHIE

1

ALMY Millie Corinne, 1967
Young children's thinking and the teaching of reading
in FROST Joe L., (rédacteur), 1967
Issues and innovations in the teaching of reading
SCOTT, FORESMAN, NEW YORK, 359P, (89-93)

2

ALMY Millie Corinne, CHITTENDEN Edward, MILLER Paula, 1966
Young children's thinking : Studies of some aspects of Piaget's theory
TEACHERS COLLEGE PRESS, NEW YORK, XII+153P, (139-140)

3

AMSTER Harriett, KEPPEL Geoffrey, 1968
Letter-sequence habits in children
in JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOUR 7, 326-332

4

ARISTIZABAL Enrique, 1938
Détermination expérimentale du vocabulaire écrit pour servir de base à l' enseignement de l' orthographe à l' école primaire
THESE DE DOCTORAT, UNIVERSITE DE LOUVAIN

5

BAKER Katherine E., FELDMAN Herman, 1956
Threshold-luminance for recognition in relation to frequency of prior exposure
in AMERICAN JOURNAL OF PSYCHOLOGY 69, 270-280

6

BARBE Walter, GRILK Werner, 1952
Correlations between reading factors and I.Q.
in SCHOOL AND SOCIETY 76, 134-136

7

BEERS James Wheelock, 1974
First and second grade children's developing orthographic concept of tense and lax vowels
THESE DE PH.D., UNIVERSITY OF VIRGINIA, 108P

8

BERDIANSKY Betty, CRONNELL Bruce, KOEHLER John Jr., 1969
Spelling-sound relations and primary form-class descriptions for speech-comprehension vocabulary of 6-9 year-olds
SOUTHWEST REGIONAL EDUCATION LABORATORY, INGLEWOOD, CALIFORNIA

9

BERGERON Colette, 1971
Fiches de travail. Méthode dynamique de français
LES EDITIONS PROJETS INC., MONTREAL

10

BERTAUD Marcel, 1971
Introduction à la statistique 3
LES PRESSES DE L' UNIVERSITE DE MONTREAL, MONTREAL, 331P

11
BEVINGTON W. G., 1958
Effects of age at time of entrance into grades I on subsequent achievement
in THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH 4, 6-16

12
BLANCHE-BENVENISTE Claire, CHERVEL André, 1969
L' orthographe
FRANCOIS MASPERO, PARIS, 234P

13
BLIESMER Emery P., 1955
The status of research in college reading
in CAUSEY Oscar, KINGSTON Albert J., (rédacteurs), 1955
Evaluating college reading programs
(= The fourth yearbook of the Southwest Reading Conference for Colleges and Universities)
TEXAS CHRISTIAN UNIVERSITY PRESS, FORT WORTH, (28-38)

14
BLOOMFIELD Leonard, 1933
Language
HENRY HOLT AND COMPANY, NEW YORK

15
BLOOMFIELD Leonard, 1942
Linguistics and reading
in ELEMENTARY ENGLISH REVIEW 19, 125-130, 183-186

16
BOND Guy Lorraine, TINKER Miles Albert, 1957
Reading difficulties : Their diagnosis and correction
APPLETON - CENTURY - CROFTS, NEW YORK, 486P

17
BURNEY Pierre, 1959
L' orthographe
(= Que sais-je no 685)
PRESSES UNIVERSITAIRES DE FRANCE, PARIS, 128P

18
BUSSIERES Simone, 1968
Je veux lire. Méthode de lecture spontanée - 1er livret
PEDAGOGIA, MONTREAL

19
BUSSIERES Simone, 1968
Je veux lire. Méthode de lecture spontanée - 2e livret
PEDAGOGIA, MONTREAL

20
BUSSIERES Simone, 1968
Je veux lire. Méthode de lecture spontanée - 3e livret
PEDAGOGIA, MONTREAL

21
BUSSIERES Simone, 1971
Méthode de lecture spontanée : guide méthodologique
PEDAGOGIA, MONTREAL, 92P

22
BUYSE Raymond, 1935
L' expérimentation en pédagogie
EDITIONS LAMERTIN, BRUXELLES, 468P

23
CALFEE R., VENEZKY Richard Lawrence, CHAPMAN R., 1969
Pronunciation of synthetic words with predictable and unpredictable letter-sound correspondences
(= Tech Report No. 71)
WISCONSIN RESEARCH AND DEVELOPMENT CENTER FOR COGNITIVE LEARNING, UNIVERSITY OF WISCONSIN, MADISON

24
CATACH Nina, 1973
Table ronde sur la structure de l' orthographe française : compte rendu
in CATACH Nina, (rédacteur), 1973
L' orthographe
(= Langue française 20)
LAROUSSE, PARIS, 128P, (6-10)
Notions actuelles d' histoire de l' orthographe
in CATACH Nina, (rédacteur), 1973
L' orthographe
(= Langue française 20)
LAROUSSE, PARIS, 128P, (11-18)

25
CATACH Nina, 1973
Que faut-il entendre par système graphique du français ?
in CATACH Nina, (rédacteur), 1973
L' orthographe
(= Langue française 20)
LAROUSSE, PARIS, 128P, (30-44)

26
CATACH Nina, (rédacteur), 1973
L' orthographe
(= Langue française 20)
LAROUSSE, PARIS, 128P

27
CATTELL J. M., 1885
Ueber die Zeit der Erkennung und Benennung von Schriftzeichen
Bildern und Farben
in PHILOSOPHISCHE STUDIEN 2, 635-650
in GIBSON Eleanor Jack, 1970
The ontology of reading
in AMERICAN PSYCHOLOGIST 25, 136-143

28
CHOMSKY Carol, 1970
Reading, writing and phonology
in HARVARD EDUCATIONAL REVIEW 40, 287-359

29
CHOMSKY Noam, HALLE Morrin, 1968
The sound pattern of English
HARPER AND ROW, NEW YORK, XIV+470P

30
COSTERMANS J., 1967
Analyse factorielle de la perception de mots chez des sujets à
audition normale
in L' ANNEE PSYCHOLOGIQUE 67:2, 393-408

31
CRONNELL Bruce, 1971
Spelling to sound correspondences for reading vs sound to spelling
correspondences for writing
(= Professional paper 12)
SOUTHWEST REGIONAL LABORATORY FOR EDUCATIONAL RESEARCH AND DEVELOPMENT,
INGLEWOOD, CALIFORNIA

32
COUSINET R., 19..
L' enseignement de la grammaire
Cahiers de pédagogie et de psychologie de l' enfant
DELACHAUX ET NIESTLE, NEUCHATEL

33
DAILEY Charles Merrill, 1975
A comparison of the spelling performance of children taught traditional
orthography with that of children taught the initial teaching
alphabet (i-t-a) in the Crawfordsville community schools, grade three
through six
THESE DE ED.D., INDIANA UNIVERSITY, 147P

34
DAYHAW Lawrence-T., 1969
Manuel de statistique
EDITIONS DE L' UNIVERSITE D' OTTAWA, OTTAWA, 550P, (2e EDITION)

35
DOUCET P. E., 1971
Comparaison de deux méthodes de cueillette de vocabulaire :
l' association libre et la rédaction spontanée au niveau de la
cinquième année du cours primaire
THESE DE LICENCE EN PEDAGOGIE, UNIVERSITE DE SHERBROOKE

36
DOWNING John A., 1969
How children think about reading
DISTINGUISHED READER'S ADDRESS, ANNUAL CONVENTION OF THE INTERNATIONAL
READING ASSOCIATION, KANSAS CITY, MAY, 1969

37
DUBOIS Félix, 1939
Echelle objective d' orthographe usuelle française
THESE DE DOCTORAT, UNIVERSITE DE LOUVAIN

38

DUBOIS Jean, GIACOMO Mathée, GUESPIN Louis,
MARCELLESI Christiane, MARCELLESI Jean-Baptiste,
MEVEL Jean-Pierre, 1973
Dictionnaire de linguistique
LAROUSSE, PARIS, 516P

39

DURREL Donald D., 1933
The influence of reading ability on intelligence measures
in JOURNAL OF EDUCATIONAL PSYCHOLOGY 24, 412-416

40

EL WAKIL M. A., 1969
L' orthographe d' usage. Contribution expérimentale à la
rationalisation de son enseignement
ROUSSEAU, GENEVE

41

ELKIND David, HORN John, SCHNEIDER Gerrie, 1965
Modified word recognition, reading achievement and perception
decentralization
in JOURNAL OF GENETIC PSYCHOLOGY 107, 235-251

42

FINK Richard S., 1974
Orthography and the perception of stops after S
in LANGUAGE AND SPEECH 17:2, 152-159

43

GAGNE Robert Mills, 1970
The conditions of learning
HOLT, RINEHART AND WINSTON, NEW YORK, (2E EDITION), VIII+407P

44

GIBSON Eleanor Jack, 1970
The ontogeny of reading
in AMERICAN PSYCHOLOGIST 25, 136-143

45

GIBSON Eleanor Jack, 1971
Perceptual learning and the theory of word perception
in COGNITIVE PSYCHOLOGY 2, 351-368

46

GIBSON Eleanor Jack, OSSER Harry, PICK Anne Danielson, 1963
A study of the development of grapheme-phoneme correspondences
in JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOUR 2, 142-146

47

GIBSON Eleanor Jack, PICK Anne Danielson, OSSER Harry,
HAMMOND Marcia, 1962
The role of grapheme-phoneme correspondences in the perception of words
in AMERICAN JOURNAL OF PSYCHOLOGY 75, 554-570

48

GIBSON Eleanor Jack, SHURCLIFF Arthur, YONAS Albert, 1970
Utilization of spelling patterns by deaf and hearing subjects
in LEVIN Harry, WILLIAMS Joanna P., (rédacteurs), 1970
Basic studies in reading
BASIC BOOKS, NEW YORK, XII+293P, (57-73)

49

GILLOOLY William Bernard, 1973
The influence of writing-system characteristics on learning to read
in READING RESEARCH QUARTERLY 8:2, 167-199

50

GOODMAN Kenneth S., 1967
La lecture, un jeu de tâtonnement psycholinguistique
in JOURNAL OF THE READING SPECIALIST, MAY, 1967
TRADUCTION FRANCAISE DE LA DIRECTION GENERALE DE L' ENSEIGNEMENT
ELEMANTAIRE ET SECONDAIRE, MINISTERE DE L' EDUCATION DU QUEBEC,
QUEBEC

51

GOODMAN Kenneth S., 1971
Reading : A psycholinguistic guessing game
in SINGER Harry, RUDDELL Robert B., (rédacteurs), 1971
Theoretical models and processes of reading
INTERNATIONAL READING ASSOCIATION, NEWARK, (259-272)

52

GREVISSE Maurice, 1975
Le bon usage : grammaire française avec des remarques sur la langue
française d' aujourd'hui
EDITIONS J. DUCULOT, S.A., GEMBLOUX, 1322P

53
GUINEBRETIERE Marie-Antoinette, 1968
Bonjour. Méthode dynamique de lecture
EDUCATION NOUVELLE, MONTREAL, 47P

54
GUINEBRETIERE Marie-Antoinette, BERGERON Colette, 1971
La méthode dynamique de lecture et de français à l' élémentaire.
Guide général
EDUCATION NOUVELLE, MONTREAL, 46P

55
GUION Jean, 1973
Nos enfants et l' orthographe
LE CENTURION, PARIS, 119P

56
GUION Jean, 1974
L' institution orthographe : à quoi sert l' orthographe, à quoi sert
son enseignement ?
LE CENTURION / SCIENCES HUMAINES, PARIS, 203P

57
HAAS William, 1970
Phono-graphic translation
MANCHESTER UNIVERSITY PRESS, MANCHESTER, 94P

58
HAGE Dean S., STROUD James B., 1959
Reading proficiency and intelligence scores, verbal and non-verbal
in JOURNAL OF EDUCATIONAL RESEARCH 52, 258-262

59
HOFFMAN Jakob, 1927
Experimentell - psychologische Untersuchungen ueber Leselerstungen
von Schulkindern
in ARCH. GESAM PSYCHOLOGIE 58, 325-388
in GIBSON Eleanor Jack, PICK Anne Danielson, OSSER Harry,
HAMMOND Marcia, 1962
The role of grapheme - phoneme correspondence in the
perception of words
in AMERICAN JOURNAL OF PSYCHOLOGY 75, 554-570

60
HOWES Davis H., SOLOMON Richard L., 1951
Visual duration threshold as a function of word probability
in JOURNAL OF EXPERIMENTAL PSYCHOLOGY 41, 401-410

61
HOYT Jeanne S., BLACKMORE Dorothy S., 1960
Fifty seventh graders : A comparison of their reading achievement
and expected achievement in grades one through seven
in JOURNAL OF EDUCATIONAL RESEARCH 53, 163-171

62
KING Ellison P., JENKINS James J., 1954
The durational threshold of visual recognition as a function of
word frequency
in AMERICAN JOURNAL OF PSYCHOLOGY 67, 700-703

63
LAMARCHE Rolande, (à paraître)
Etude du code phono-graphique d' adultes français
UNIVERSITE DE MONTREAL

64
LAMBERT Jean, 1947
Contribution à la didactique expérimentale de l' orthographe d' usage :
l' analyse des fautes
THESE DE DOCTORAT, UNIVERSITE DE LOUVAIN

65
LE CONSEIL DE DEVELOPPEMENT SOCIAL DE MONTREAL (C.D.S.M.), 1970
L' éducation dans les milieux défavorisés
in BELANGER Pierre W., ROCHER Guy, (rédacteurs), 1970
Ecole et société au Québec : éléments d' une sociologie de
éducation
HMH, MONTREAL, 465P, (329-359)

66
LEFTON Lester A., SPRAGINS Anne B., BYRNES John, 1973
English orthography : Relation to reading experience
in BULLETIN OF THE PSYCHONOMIC SOCIETY 2(5-A, 281-282

67

LENNON Roger T., 1950
The relation between intelligence and achievement test results for a group of communities
in JOURNAL OF EDUCATIONAL PSYCHOLOGY 41, 301-308

68

LIBERMAN Isabelle Y., SHANKWEILER Donald, CARTER Bonnie, FISCHER F. William, 1972
Reading and the awareness of linguistic segments
in HASKINS LABORATORIES STATUS REPORT ON SPEECH RESEARCH 31/32, 149-157

69

LOWRY Ellsworth, 1932
Increasing the I.Q.
in SCHOOL AND SOCIETY 35, 179-180

70

LUSSIER Colette, PICARD Raymonde, POULIOT Monique, 1977
Mise en marche du processus de l' apprentissage de la lecture : méthode dynamique de lecture
LES EDITIONS PROJETS INC., MONTREAL

71

MARTINET André, 1960
Eléments de linguistique générale
, ARMAND COLIN, PARIS, 221P

72

MASON Mildred, 1975
Reading ability and letter search time : Effects of orthographic structure defined by single-letter positional frequency
in JOURNAL OF EXPERIMENTAL PSYCHOLOGY : GENERAL, 104:2, 146-166

73

MASSARENTI Léonard, 1970
Psychopédagogie des moins doués
RALLE, SUISSE, 256P

74

MCGINNIES Elliott M., COMER Patrick B., LACEY Oliver L., 1952
Visual recognition threshold as a function of word length and word frequency
in JOURNAL OF EXPERIMENTAL PSYCHOLOGY 44, 65-69

75

MCNEMAR Quinn, 1969
Psychological statistics
JOHN WILEY AND SONS INC., NEW YORK, VII+529P

76

MINKLER Rachel S., 1959
Developmental patterns in the fundamental skills of the elementary school programme
in ONTARIO JOURNAL OF EDUCATIONAL RESEARCH 2, 45-50

77

MORF Albert, 1966
Elaboration d' une méthode fonctionnelle d' apprentissage de la lecture
RAPPORT PRESENTE AU MINISTERE DE L' EDUCATION, JUIN 1966

78

MOSCOVICI S., HUMBERT T. C., 1960
Usage et disponibilité comme facteurs déterminant la durée de stimuli verbaux

79

NEISSER Ulric, 1967
Cognitive psychology
APPLETON - CENTURY - CROFTS, NEW YORK, XI+351P

80

PEIXOTTO Helen E., 1948
The recognitive value of three hundred nonsense syllables
in AMERICAN JOURNAL OF PSYCHOLOGY 61, 352-360

81

PELLET Mlle, 1926
A quel âge sont acquises les différentes règles de grammaire ?
(= Bulletin no 212-213)
SOCIETE ALFRED BINET ET THEODORE SIMON, LYON

82

PETTY Walter T., MURPHY J. Brian, MOHAN Madan, 1974
Spelling achievement and the initial teaching alphabet : Analysis of errors
in ELEMENTARY SCHOOL JOURNAL 74:5, 309-313

83

PIACERE Jean, 1974
Introduction au fonctionnement psychologique et linguistique de l' orthographe française
in DOCUMENTS ET RECHERCHES, LETTRES, 4, 5-9

84

PIAGET Jean, 1963
The psychology of intelligence
LITTLEFIELD, ADAMS, NEW YORK, 182P

85

PIRENNE Albert, 1949
Programme d' orthographe d' usage pour les écoles primaires
LA PROCURE, NAMUR

86

PORQUET André, 1973
Réforme orthographique et graphémologie
in CATACH Nina, (rédacteur), 1973
L' orthographe
(= Langue française 20)
LAROUSSE, PARIS, 128P, (45-51)

87

POSTMAN Leo, CONGER B., 1954
Verbal habits and the visual recognition of words
in SCIENCE 19, 671-673

88

POSTMAN Leo, ROSENZWEIG Mark R., 1956
Practice and transfer in the visual and auditory recognition of verbal stimuli
in AMERICAN JOURNAL OF PSYCHOLOGY 69, 209-226

89

PREFONTAINE Robert R., PREFONTAINE Gisèle, 1963
Vocabulaire oral des enfants de 5 à 8 ans au Canada français
BEAUCHEMIN, MONTREAL, 149P

90

PRESTON Ralph Claustus, 1953
Comparison of word-recognition skill in German and American children
in ELEMENTARY SCHOOL JOURNAL 53, 443-446

91

PRESTON Ralph Claustus, 1963
Issues raised by the Washoden - Philadelphia reading study
in COMPARATIVE EDUCATION REVIEW 7, 61-65

92

RAMBAUD Honorat, 1578
La déclaration des abus que l' on commet en escrivant et le moyen de les éuiter et représenter nayuement les paroles : ce que jamais homme n' a faict
EDITIONS JEAN DE TOURNES, LYON, (15-78)

93

RAVEN Ronald J., SALZER Richard T., 1971
Piaget and reading instruction
in THE READING TEACHER 24, 630-640

94

REGER Roger, 1962
Brief tests of intelligence and academic achievement
in PSYCHOLOGICAL REPORT 11, 82

95

REICHENBACH Daniel, 1956
Essai de programme d' orthographe d' usage pour les 8e et 9e années de l' école primaire
EDUCATEUR, 30, MONTREUX

96

ROBERTSON Douglas J., TREPPER Terry S., 1974
The effect of I.T.A. on the reading achievement of Mexican-American children
in READING WORLD 15:2, 132-139

97

ROGERS T. S., 1967
Linguistic consideration in the design of the Stanford computer - based on a curriculum in initial reading
(= Technical Report No. 111)
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES, STANFORD UNIVERSITY, STANFORD

98
ROLLER Samuel, 1943
Les enseignements d' une dictée
Cahiers de pédagogie et de psychologie de l' enfant
DELACHAUX ET NIESTLE, NEUCHATEL

99
ROSINSKI Richard R., WHEELER Kirk E., 1972
Children's use of orthographic structure in word discrimination
in PSYCHONOMIC SCIENCE 26, 97-98

100
ROSSI Peter H., 1968
Social factors in academic achievement : A brief review
in HALSEY A. H., FLOUD Jean, ANDERSON Arnold C.,
(rédacteurs), 1968
Education, economy and society
THE FREE PRESS, NEW YORK ; COLLIER-MACMILLAN LIMITED, LONDON,
625P, (269-272)

101
RUDDEL Robert B., 1968
The relation of regularity of grapheme-phoneme correspondences and
language structure to achievement in first-grade reading
in GOODMAN Kenneth S., (rédacteur), 1968
The psycholinguistic nature of the reading process
WAYNE STATE UNIVERSITY PRESS, DETROIT, 347P, (255-270)

102
SAMUELS Stanley Jay, 1969
Cross-national studies in reading : The relationship between the
sound-letter correspondence in language and reading achievement
in FIGUREL J. Allen, (rédacteur), 1969
Reading and realism
INTERNATIONAL READING ASSOCIATION, NEWARK, DELAWARE, XII+890P,
(846-853)

103
SANTA Carol Minnick, 1976
Spelling patterns as units of word perception for adults and children
reading on first, second and fifth grade reading levels
THESE DE PH.D., TEMPLE UNIVERSITY, 140P

104
SAPIR Edward, 1921
Language : An introduction to the study of speech
HARCOURT-BRACE COMPANY, NEW YORK

105
SAUSSURE Ferdinand De, 1916
Cours de linguistique générale
EDITIONS CHARLES BALLY - ALBERT SECHEHAYE, LAUSANNE, PAYOT

106
SCHWARTZ Sybil, 1975
A developmental study of children's ability to acquire knowledge of
spelling pattern
THESE DE PH.D., MCGILL UNIVERSITY, 130P
Le genre des adjectifs en français
in CATACH Nina, (rédacteur), 1973
L' orthographe
(= Langue française 20)
LAROUSSE, PARIS, 128P, (52-74)

107
SEXTON Patricia C., 1965
Education and income : Inequalities in our public schools
THE VIKING PRESS, NEW YORK, (2e EDITION)

108
SHANNON Claude E., 1951
Prediction and entropy of printed English
in BELL SYSTEMS TECHNICAL JOURNAL 30, 50-64

109
SMITH Frank, 1971
Understanding reading : A psycholinguistic analysis of reading and
learning to read
HOLT, RINEHART AND WINSTON, NEW YORK, 260P

110
SMITH Frank, 1973
Psycholinguistic insights into language learning
in ROUDEAU Guy, (rédacteur), 1973
Contributions canadiennes à la linguistique appliquée
CENTRE EDUCATIF ET CULTUREL, MONTREAL, 272P, (163-175)

SMITH Wendell I., ROHRMAN Nicholas L., 1968
Human learning
MCGRAW-HILL, NEW YORK, 173P
112

SOLOMON Richard L., POSTMAN Leo, 1952
Frequency of usage as a determinant of recognition threshold for words
in JOURNAL OF EXPERIMENTAL PSYCHOLOGY 43, 195-201
113

SPIEGEL Murray R., 1961
Theory and problems of statistics
SCHAUM PUBLISHING COMPANY, NEW YORK, 359P
114

STAATS Arthur Wilbur, 1968
Learning language and cognition
HOLT, RINEHART AND WINSTON, NEW YORK, XVII+614P
115

TAYLOR Janet A., 1958
Meaning, frequency and visual duration threshold
in JOURNAL OF EXPERIMENTAL PSYCHOLOGY 55, 329-334
116

TERS François, 1973
L' orthographe dans son contexte sociolinguistique
in CATACH Nina, (rédacteur), 1973
L' orthographe
(= Langue française 20)
LAROUSSE, PARIS, 128P, (75-85)
117

TERS François, MAYER Georges, REICHENBACH Daniel, 1964
Vocabulaire orthographique de base
MESSEILLER, NEUCHATEL
118

TERS François, MAYER Georges, REICHENBACH Daniel, 1977
L' échelle Dubois-Buyse d' orthographe usuelle française
EDITIONS O.C.D.L., PARIS, (4e EDITION), 94P
119

THIMONNIER René, 1970
Code orthographique et grammatical
HATIER, PARIS, 320P
120

ULDALL H. J., 1944
Speech and writing
in ACTA LINGUISTICA 4, 11-16
121

VACHEK Josef, 1959
Two chapters on written English
in BRNO STUDIES IN ENGLISH 1, 7-38
Written language : general problems and problems of English
MOUTON, LA HAYE-PARIS, 80P
122

VEIRASSE D'ALLAIS Denis, 1681
Grammaire méthodique contenant en abrégé les principes de cet art et les règles les plus nécessaires de la langue française dans un ordre clair et naturel
PARIS
123

VENEZKY Richard Lawrence, 1967
English orthography : Its graphical structure and its relation to sound
in READING RESEARCH QUARTERLY 2, 75-106
124

VENEZKY Richard Lawrence, 1973
Letter-sound generalization of first, second and thrid grade Finnish children
in JOURNAL OF EDUCATIONAL PSYCHOLOGY 64:3, 288-292
125

VENEZKY Richard Lawrence, WEIR Ruth H., 1966
A study of selected spelling to sound correspondence patterns
(= Cooperative Research Project No. 3090)
UNITED STATES OFFICE OF EDUCATION - STANFORD UNIVERSITY, STANFORD
126

WALLACH Michael A., 1963
Perceptual recognition of approximations to English in relationship to spelling achievement
in JOURNAL OF EDUCATIONAL PSYCHOLOGY 54, 57-62

WILLIAMS Joanna P., 1968
Successive versus concurrent presentation of multiple grapheme-phoneme correspondences
in JOURNAL OF EDUCATIONAL PSYCHOLOGY 59:5, 309-314

La collection **Le français à l'école primaire** a pour objectif de faire connaître des documents d'intérêt varié produits dans le cadre du P.P.M.F. primaire de l'Université de Montréal. Ces documents s'adressent à toute personne intéressée à la didactique du français, langue maternelle, tant du point de vue pratique que théorique. La collection compte trois séries.

Perfectionnement des maîtres

Cette série regroupe des textes, des mises en situation et des ateliers destinés à susciter la réflexion et à motiver un renouvellement de la pratique de l'enseignement du français. Les professeurs chargés de la formation ou du perfectionnement des maîtres de français, les conseillers pédagogiques et les maîtres de français peuvent s'y référer.

Recherches et essais

Cette série présente les résultats de recherches et d'études produites par les professeurs, agents de recherche et conseillers pédagogiques reliés au P.P.M.F. primaire de l'Université de Montréal. Fruits d'une démarche expérimentale, analytique ou exploratoire, les documents de cette série abordent des aspects divers de la langue, de son fonctionnement et de son apprentissage, sous l'angle de la linguistique, de la psychologie ou de la pédagogie.

Documentation pédagogique

Cette série propose une documentation à caractère pratique issue de travaux et d'expérimentations menés dans la classe de français ou recensée à travers la littérature sur la didactique de la langue maternelle. Regroupées par dossiers, les situations d'apprentissage et les interventions pédagogiques suggérées sont directement utiles aux maîtres de français.

Collection publiée par le Centre de diffusion du Programme de perfectionnement des maîtres de français au primaire (P.P.M.F.)
Faculté des sciences de l'éducation
Université de Montréal